PREDICTION OF HOURLY SOLAR RADIATION

UC3M

Elianne Mora
Advanced Regression and Prediction

Contents

Interest

Renewable energy is the key to sustainability and one way to tackle the negative effects of climate change. It is crucial that countries effectively assess the potential of their diverse landscapes to begin their sustainability journey. Not all countries were created equal, while the US may be able to establish several different sources of renewable energy throughout the country, the same can not be said about Greece, who may be extremely limited in their approach and resources.

In this study we focus on solar energy. We previously concentrated on predicting the average daily solar radiation; however, for this portion of the project we focus on predicting the hourly solar radiation. The aim of expanding the focus of the study is to provide more detailed and reliable predictions which may foster innovation and investments in the island's solar energy development.

Understanding Solar Radiation Records

The original data set focuses on metrics recorded through the months of September-December. As it can be observed below, these months share similar patterns, except for December, which clearly shows much lower radiation records.

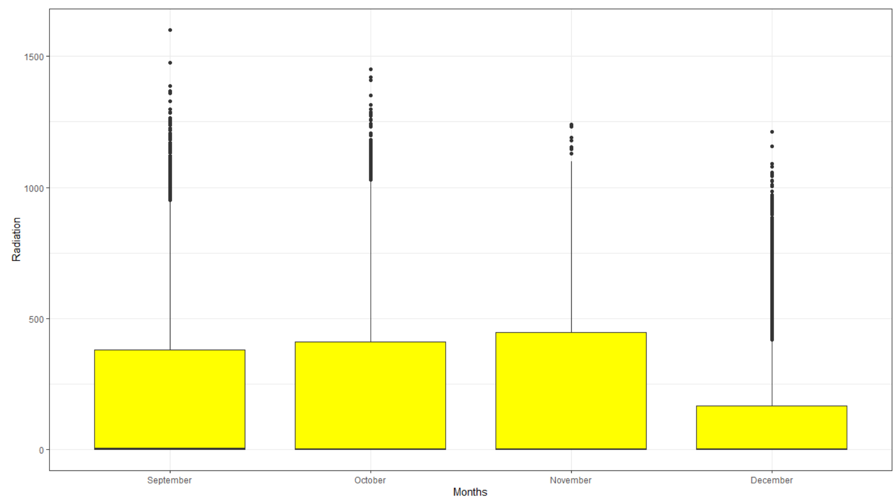

Figure 1. Solar radiation over the months.

Since the computational load of the original dataset (32,554) is excessive, we will only focus our prediction on the month of December. This month may be of special interest, given its low radiation records, when considering solar panel installation, assessment of back-up batteries and performance estimation.

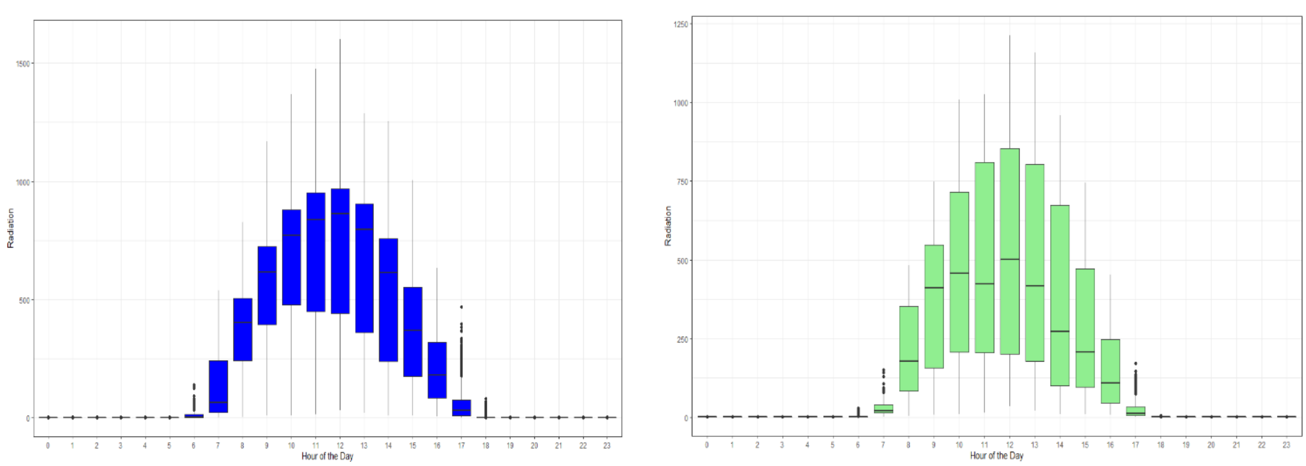

Figure 2. Hourly solar radiation for September-December (left) and hourly radiation for December only (right)

If we focus on the y-axis of figure 2, we can see that this Hawaiian island clearly experiences lower radiation in December (left) in comparison with the 3 prior months. Also, as we explore the relationship between temperature and radiation below, we notice that November and December share similar records, yet the latter experiences lower temperatures.

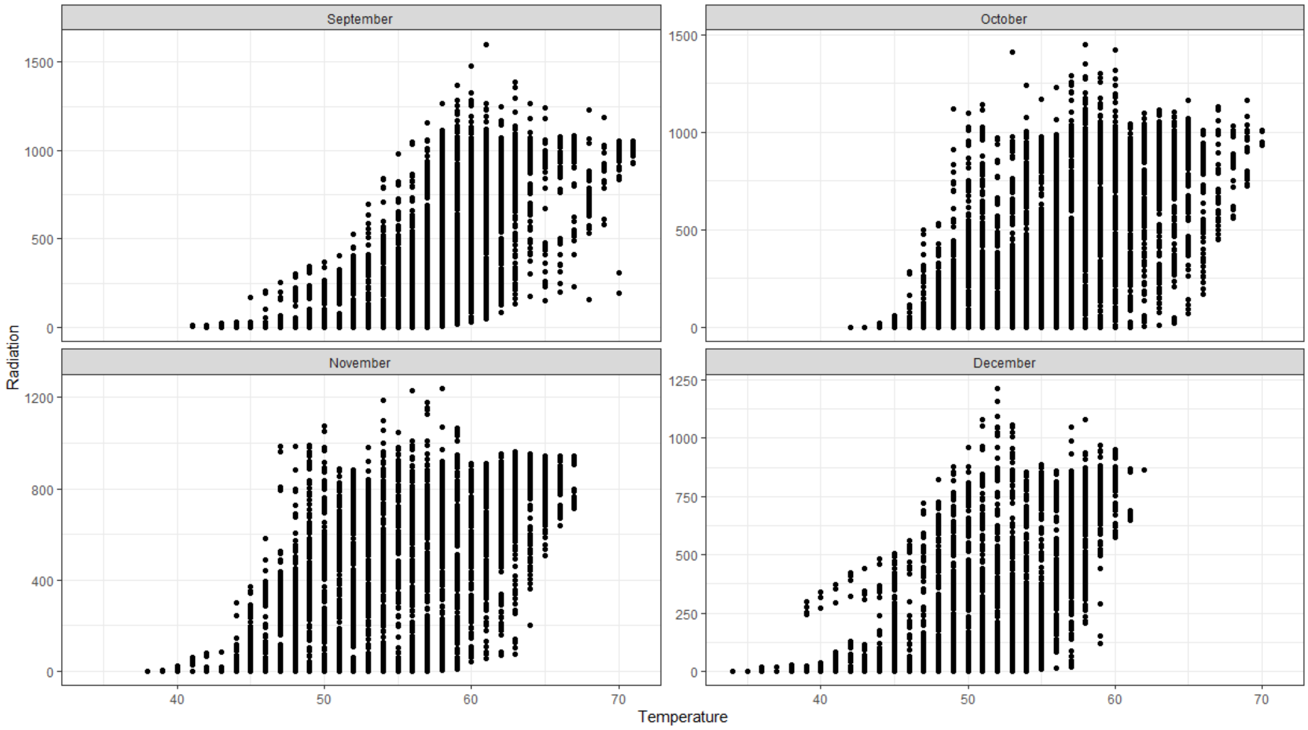

Figure 3. Radiation vs. Temperature by month.

Given this breve analysis, in the prediction of radiation we may consider covariates such as hour of the day (0-23), temperature (F), humidity (%), wind direction (degrees), and wind speed (mph).

Advanced Regression Models

We will explore different predictive models where we use a simple model such that radiation is explained by five covariates:

Radiation~Hour+Temperature+Humidity+WindDirection.Degrees.+Speed

For the sake of simplicity in the implementation of the different models, we will assume that the data is independent (rows). The data is split into a training (75%) and a testing(25%) set.

Smart Vector Regression

We fit a simple model into the SVR, defining the train control for parameter optimization using cross-validation, using different folds. The number of values to be tried to find the optimum hyperparameter C are also interchanged to grasp their effect on the model's predictive accuracy.

Linear Kernel

We begin by fitting a linear kernel, where C is held constant at 1. From the results below, the model can explain 60.86% of the variability given the fitted covariates. However, the coefficient of determination for the testing set is roughly 66.74%. It must be noted that regardless of tuning of the parameters, the coefficient of determination for the testing set remained the same.

```
Resampling results:

  RMSE      Rsquared  MAE
  142.7508  0.668631  85.72006

Tuning parameter 'C' was held constant at a value of 1
```

Figure 4. Smart vector regression with simple model, CV at 10-folds, and tune length 10.

In the plot below, the observed values appear in grey and the predicted values in blue. The model's linear fit seems poor, indicating its non-linearity.

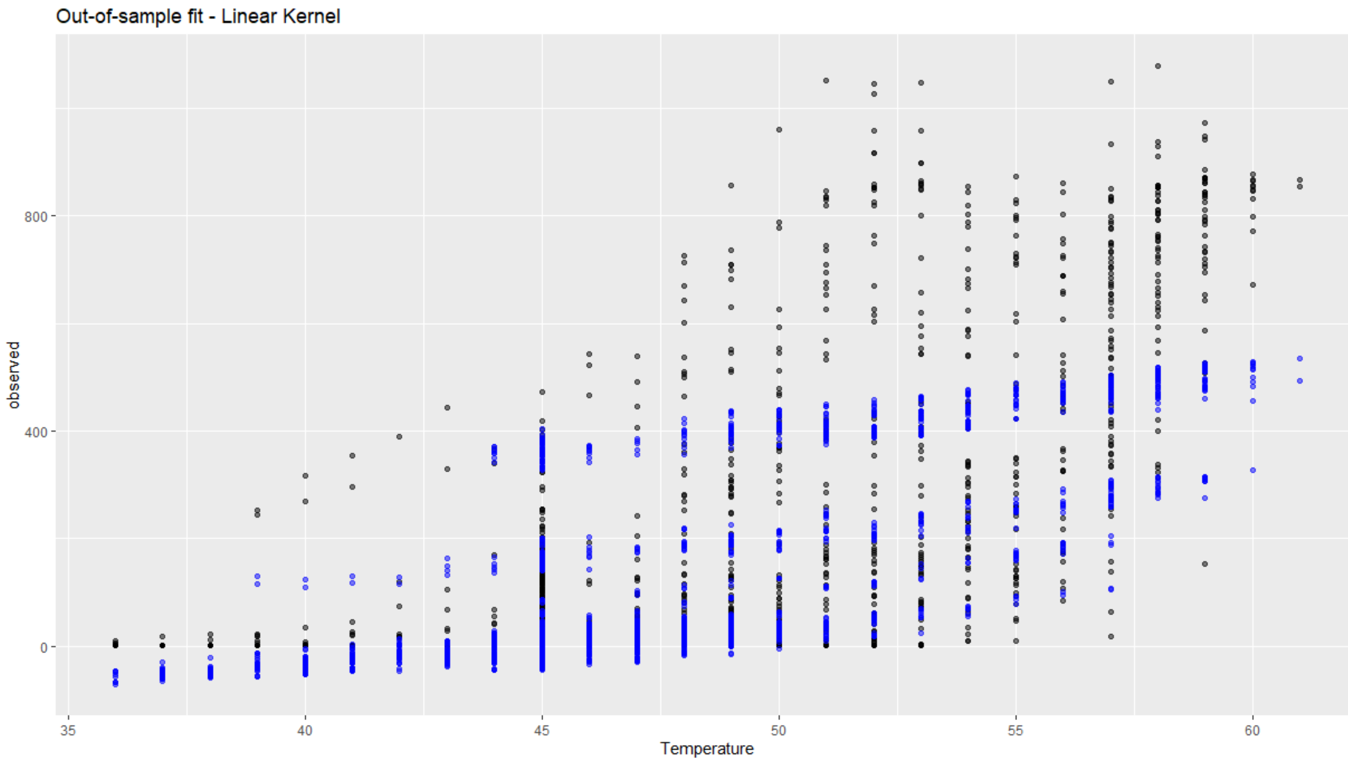

Figure 5. Smart vector regression results from linear kernel. Observed radiation values appear in grey scale and predicted values appear in blue.

Polynomial Kernel

In this instance, we use a polynomial kernel and tune the parameters as exemplified in table 1. We can observe that by increasing the number of values to be attempted to find the optimal hyperparameter C, the better the reported coefficient of determination of the testing set. Unfortunately, our computational power did not allow us to increase the parameters further under this kernel.

Number of Folds (CV)	Tune Length	Degree	Scale	C	R^2
3	2	2	0.1	0.5	0.7819
3	3	3	0.1	0.25	0.8299

Table 1. SVR - polynomial kernel implemented with 3 CV folds and a tune length of 3 for hyperparameter C.

From the plot below we can observe that the polynomial Kernel performs much better than the linear kernel.

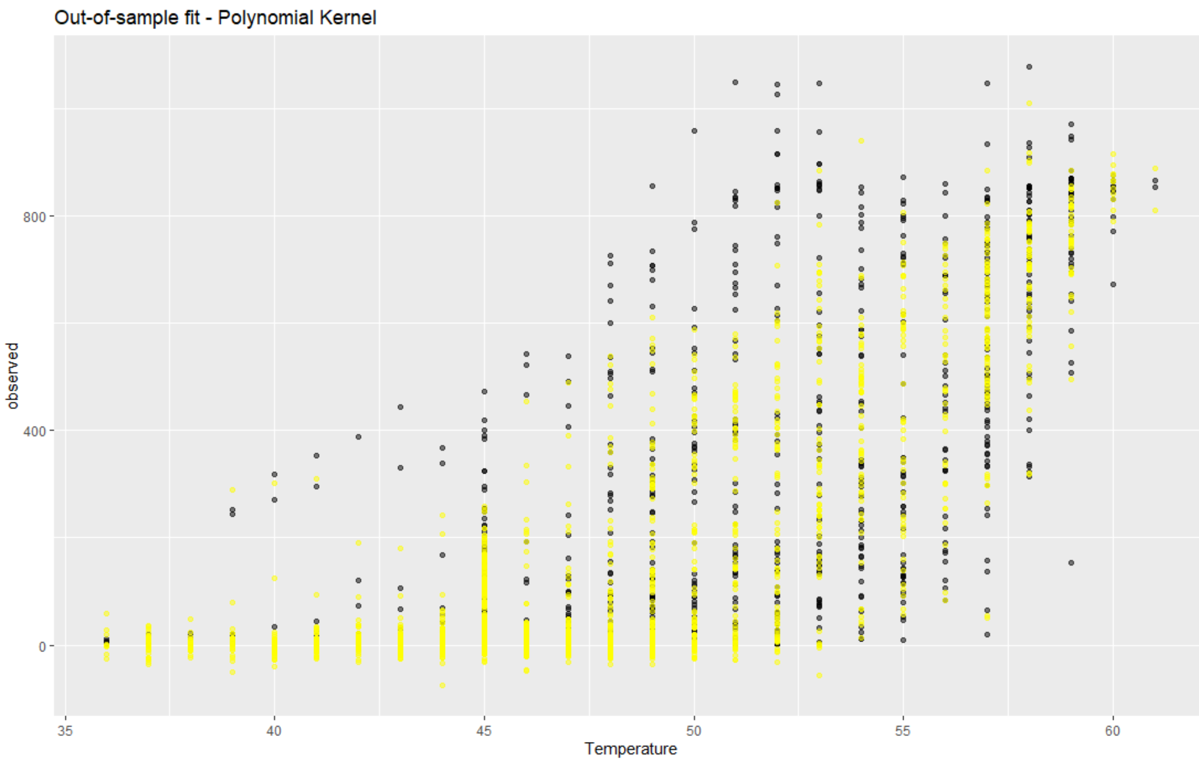

Figure 6. SMR polynomial model with 3-folds CV and 3 tune length.

Gaussian Kernel

We fit another set of SVR models using a Gaussian Kernel holding the same parameters. From the results below, the optimum value of C was 128, where the model explains 84.08% of the variability of the data. The $\mathrm{R}^2 = 84.08\%$ obtained from the testing set is slightly higher than the one obtained with the polynomial kernel.

Number of Folds (CV)	Tune Length	Sigma	R^2	C
2	2	0.0223	0.8009	0.5
3	3	0.0230	0.8087	1
4	4	0.0228	0.8146	2
6	6	0.0222	0.8279	8
10	10	0.0224	0.8408	128

Table 2. SVR - Gaussian Kernel results with different tuning of parameters.

```
Resampling results across tuning parameters:

  C        RMSE       Rsquared   MAE
    0.25   107.95042  0.8049943  58.29861
    0.50   105.79650  0.8112955  56.09630
    1.00   104.17659  0.8163192  53.86205
    2.00   102.63575  0.8213914  52.27312
    4.00   100.99983  0.8268060  51.21750
    8.00    99.63376  0.8313633  50.29269
   16.00    98.66080  0.8345011  49.53302
   32.00    97.78833  0.8373231  48.90536
   64.00    97.33899  0.8386616  48.81456
  128.00    96.76572  0.8405312  48.70135

Tuning parameter 'sigma' was held constant at a value of 0.02247058
RMSE was used to select the optimal model using the smallest value.
The final values used for the model were sigma = 0.02247058 and C = 128.
```

Figure 7. Smart vector regression with Gaussian Kernel, CV at 10-folds, and tune length 10.

The plot below shows that this model fits a bit better than the polynomial kernel, as there is a slight upward shift that accounts for some extreme values.

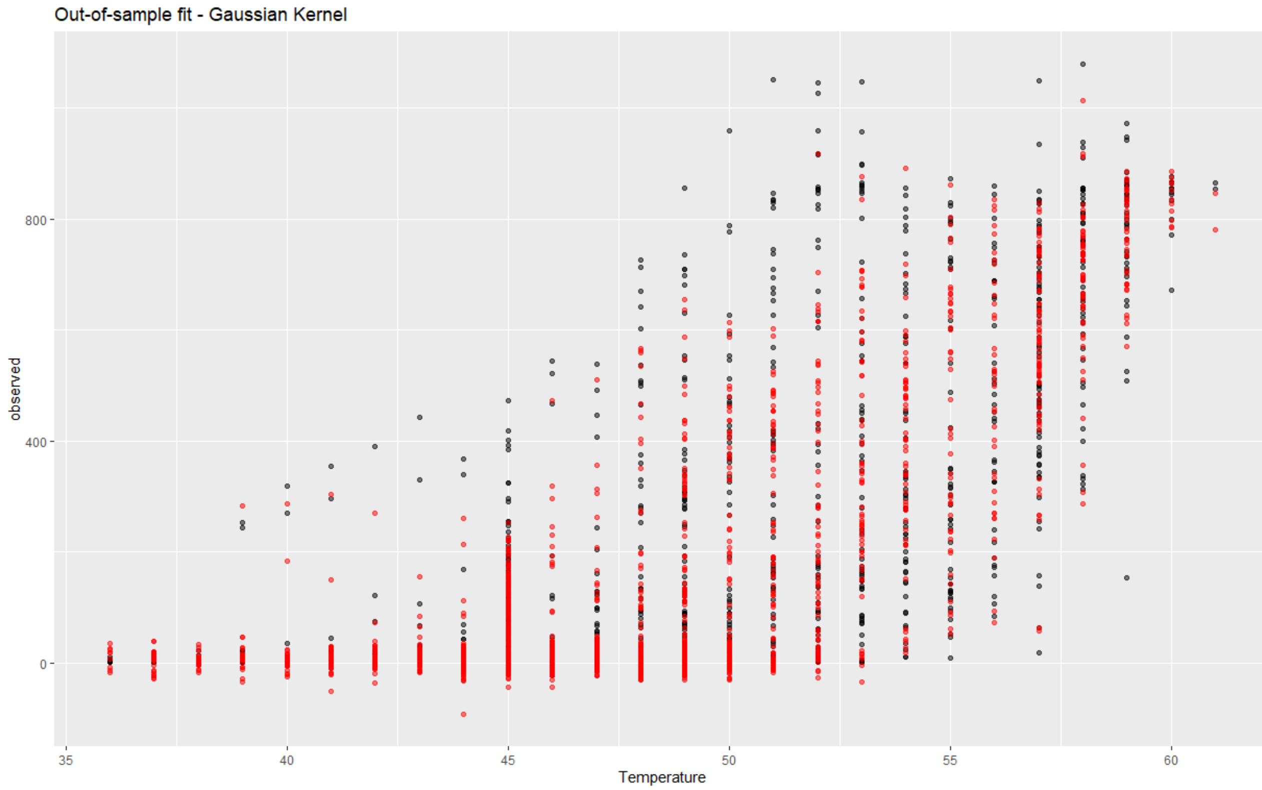

Figure 8. Smart vector regression results from Gaussian kernel. Observed radiation values appear in grey scale and predicted values appear in blue.

Random Forest

We now implement different decision trees, starting with Random Forest. The results are as follows:

Number of Folds (CV)	Tune Length	Number of Trees	R^2
3	3	100	0.8934
6	6	100	0.8975
3	2	200	0.8951
4	4	300	0.8982

The final model's coefficient of determination on the testing set is significantly higher than the previous SVR models attempted. The plot below shows the covariates in order, from most useful in predicting radiation to least useful.

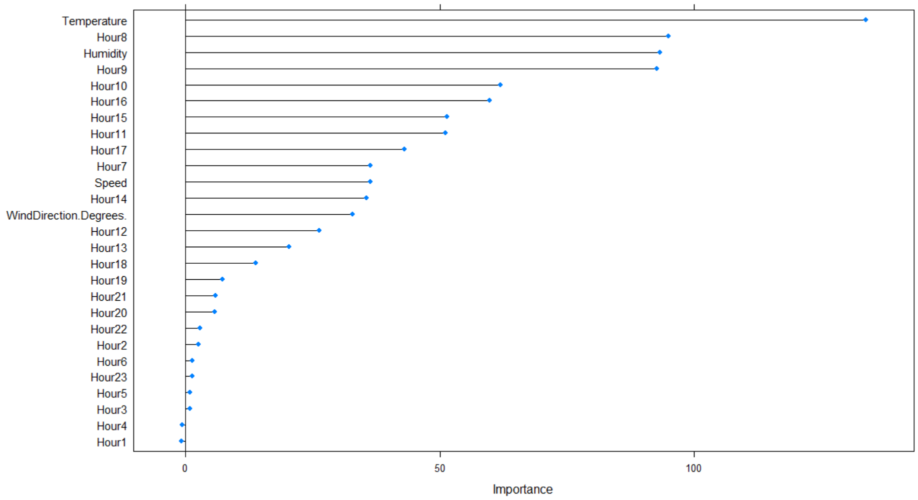

Figure 9. RF model output ighlighting covariates' usefulness in predicting solar radiation.

Plotting the model below, a slight downward move can be noticed when compared to the SVR Gaussian kernel; however, the reported R^2 for the testing set is higher in the present model.

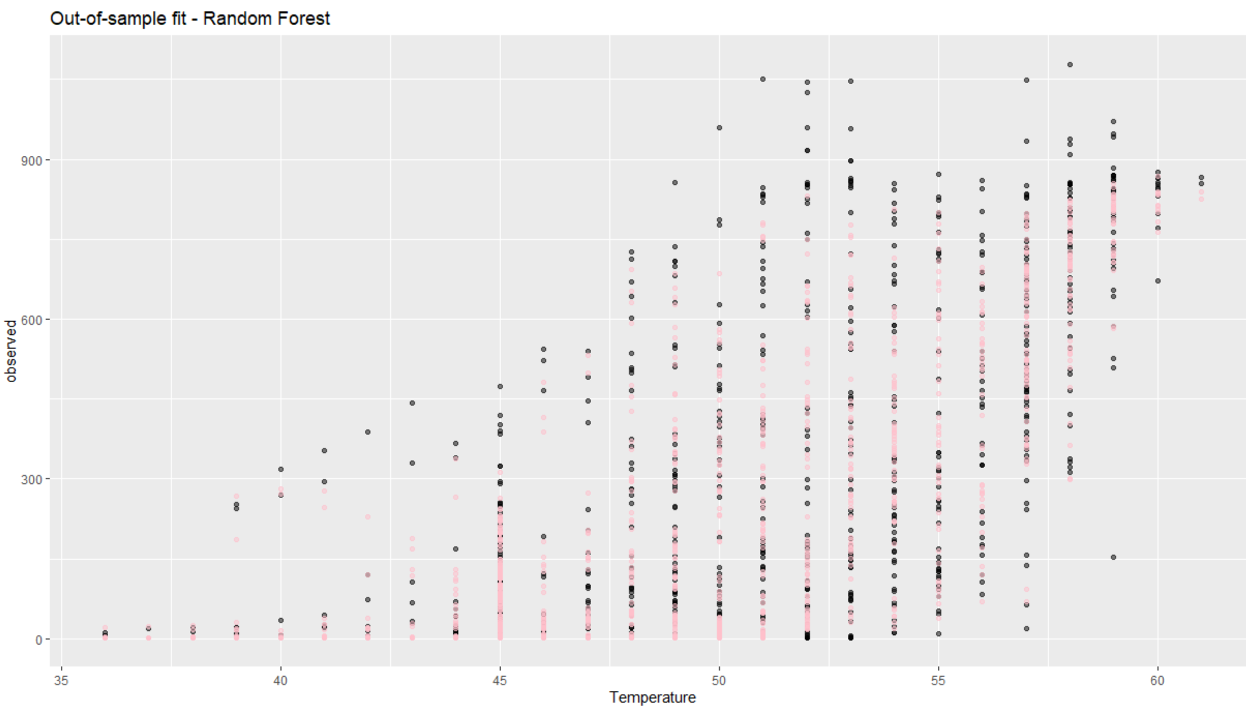

Figure 10. RF final model with 4-folds CV, 4 tune length, and 300 trees.

Gradient Boosting

Lastly, we implement a decision tree where we use gradient boosting and different parameters. The results below show that as we increase the parameters, the reported R^2 also increases.

Number of Folds (CV)	Tune Length	R^2
3	3	0.8316
6	6	0.8765
10	10	0.8842

Table 3. Gradient boosting results from tuning parameters.

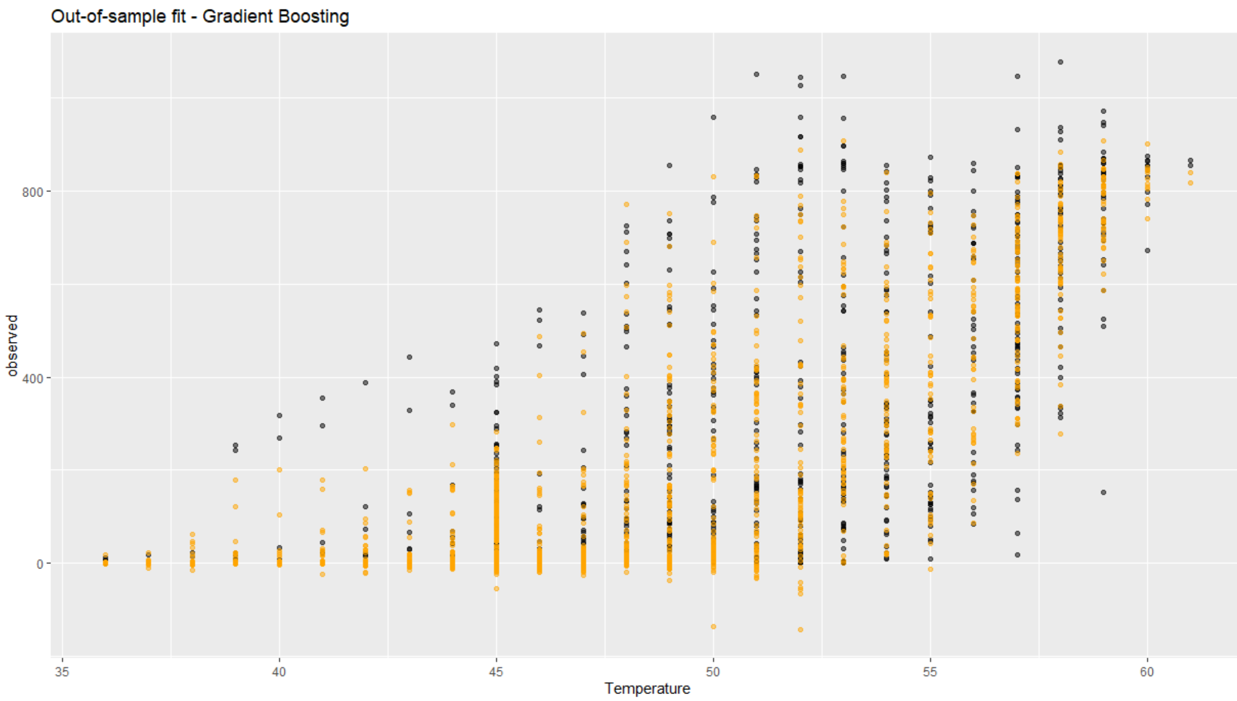

Figure 11. Final model with gradient boosting, 10-folds CV, 10 tune length.

Neural Networks

Lastly, due to computational limitations, we fit a quite simple neural network model. The results below show that the model does not possess strong predictive capabilities in comparison with the previous models.

```
Neural Network

6025 samples
   5 predictor

Pre-processing: centered (27), scaled (27)
Resampling: Cross-Validated (5 fold)
Summary of sample sizes: 4821, 4820, 4819, 4821, 4819
Resampling results across tuning parameters:

  size  decay  RMSE      Rsquared     MAE
   5    0.01   278.9035  0.076535226  138.224
   5    0.02   278.9035  0.113841090  138.224
   5    0.03   278.9035  0.051721555  138.224
  10    0.01   278.9035  0.001201877  138.224
  10    0.02   278.9035  0.009298136  138.224
  10    0.03   278.9035  0.278194886  138.224

RMSE was used to select the optimal model using the smallest value.
The final values used for the model were size = 10 and decay = 0.01.
```

Conclusion

The model that would offer the best value to investors contemplating whether to set up solar energy in the island would be the Random Forest model, with 4-folds CV and tune length, and 300 trees. It was able to explain roughly 90% of the hourly variability of the data, which is significant.

The present work highlights the importance of machine learning tools in business decisions and resource management. Some of the limitations of the study arise from the high computational expense of implementing the different models and their respective tuning of parameters.