SOLAR ENERGY FORECASTING WITH MACHINE LEARNING APPROACHES

MD Sabbir Hossain[1] and M. Siyamalan[2]
Department of Computer Science, Faculty of Science, University of Jaffna
[1]falaksabbir3@gmail.com, [2]siyam@univ.jfn.ac.lk

Abstract

Solar power has been used as an energy source by electric utility companies. These companies must forecast accurate production of solar energy to determine fuel needs. Machine Learning plays an important role for this forecasting. In the last decade many research done for forecasting solar energy using various machine learning techniques. In this work, we compare three machine learning approaches such as XGBoost, Random Forest and Linear Regression for solar energy prediction, where, XGBoost and Random Forest are non-linear algorithms and linear regression is a linear one. We show that XGBoost performs better than others.

Dataset and Experimental Settings

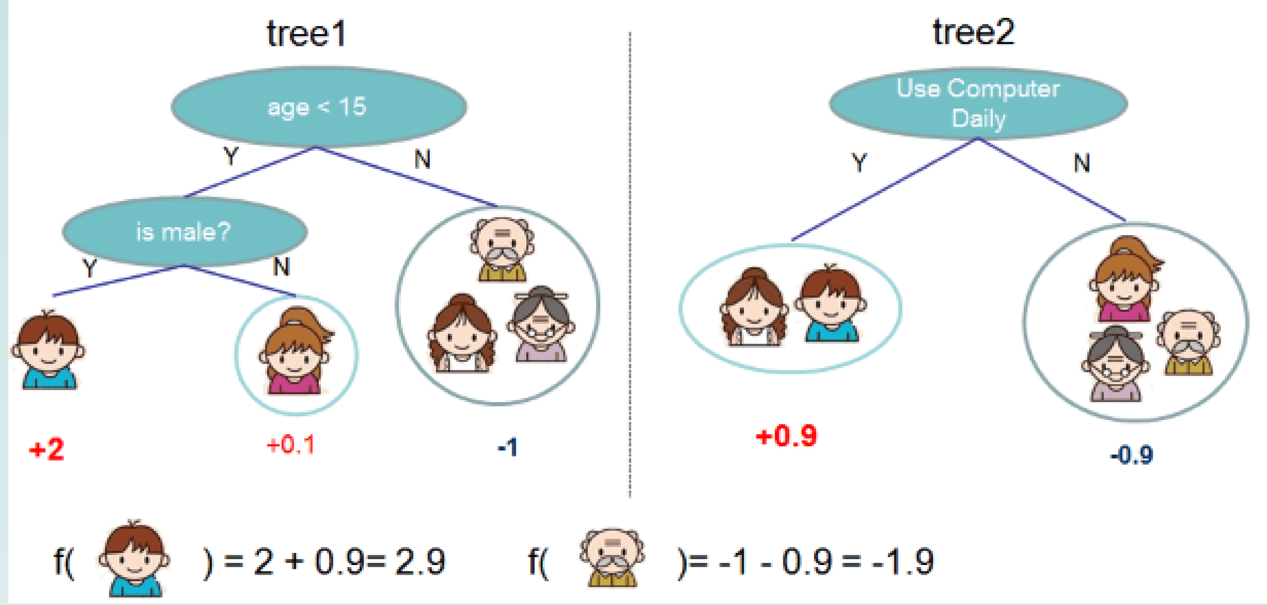
This public dataset is obtained from University of Illinois in Urbana Champaign site and detailed by National Oceanographic and Atmospheric Administration (NOAA). This is an hourly based data contains 165,803 instances. This table shows Weather Features from obtained dataset.

Weather Features	Unit
Cloud Coverage	% range
Visibility	Miles
Temperature	^{0}C
Dew Point	^{0}C
Relative Humidity	%
Wind Speed	Mph
Station Pressure	inchHg
Altimeter	inchHg

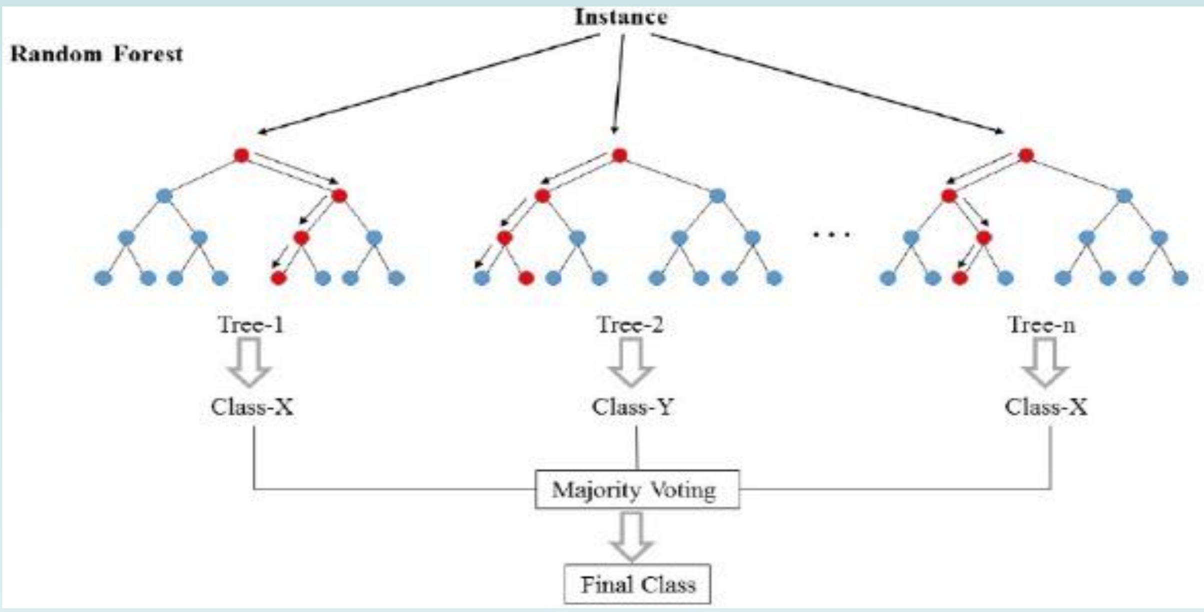
We split the dataset into training and testing set as 77% and 33%. We used Root Mean Squared Error (RMSE) as the evaluation measure.

Methodology

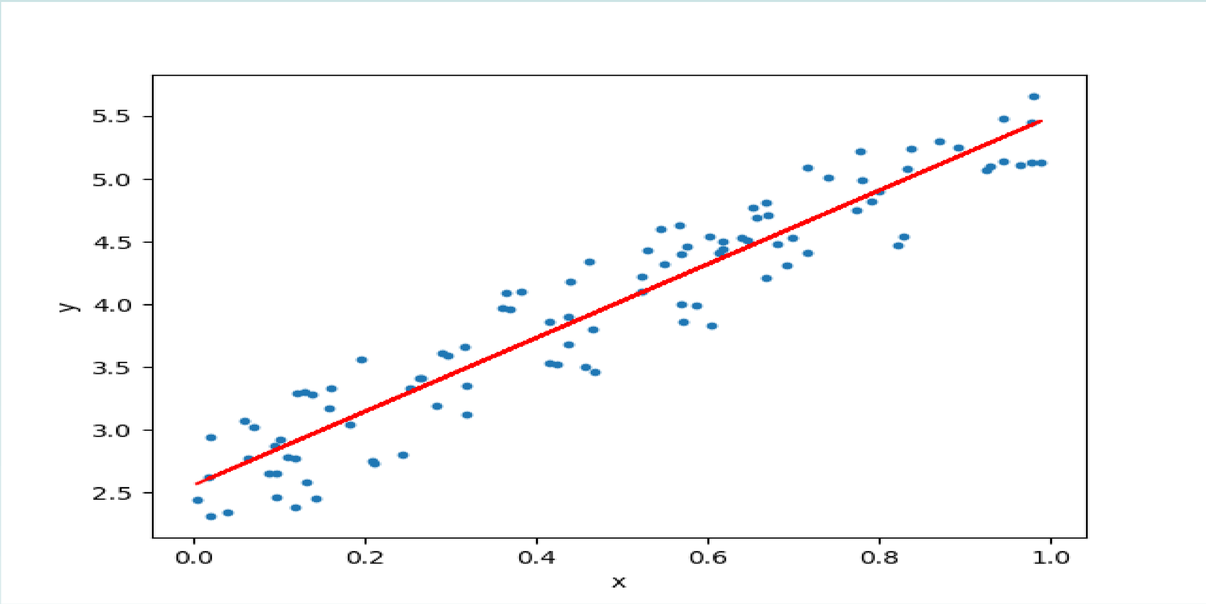
XGBoost and Random Forests are ensemble tree learners. In both cases the final prediction can be obtained by combining the predictions of individual trees. However, they differ from each other by the way the trees are learned. In XGBoost each tree is learnt in an additive manner. At each iteration a tree is found in a greedy way and added to the existing model. On the other hand, in Random Forests each tree is learnt independently from others using a subset of the entire training set. In XGBoost each tree is a weak-learner, but in Random Forests each one is a strong model. XGBoost and Random Forests are non-linear algorithms, on the other hand, Linear Regression is a linear one, Linear Regression models the relationship between a the features and the outcomes.



XGBoost



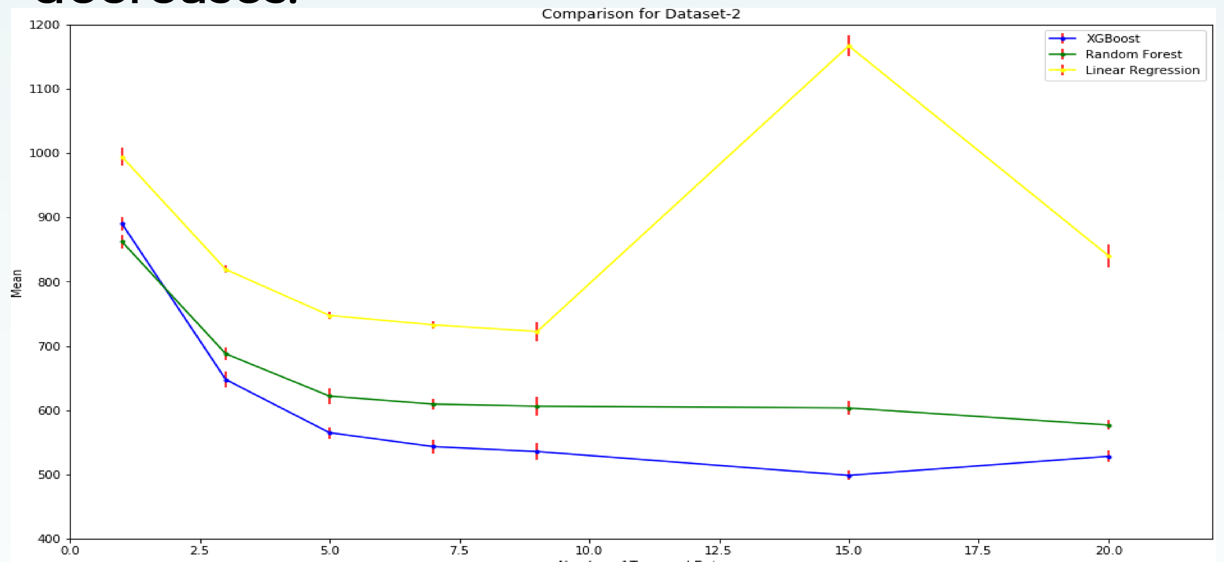
Random Forest



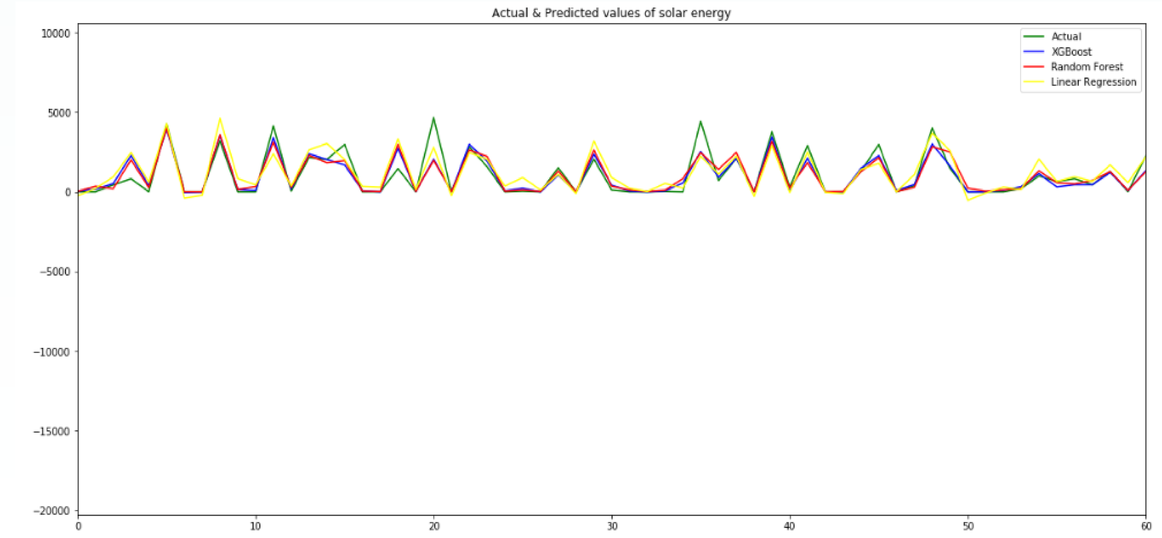
Linear Regression

Results

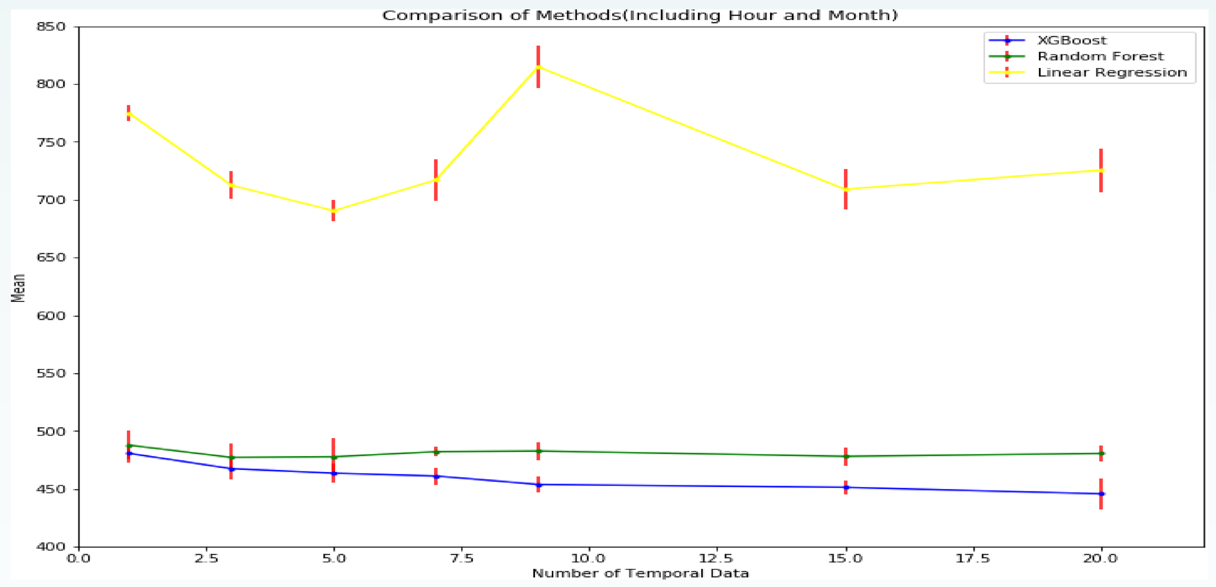
The following figure shows the RMSE values for each method. We compared all the methods with and without temporal data. We found that with temporal data all methods perform better. When the number of temporal data increases, the value of RMSE decreases.



The following figure shows the difference between actual solar energy and predicted solar energy.



There is a big difference between starting and ending index for each month and hour. To improve the performance, we removed this difference by mapping month and hour. The following figure shows the RMSE value for each method with mapped hour and month.



The following table shows the RMSE value for each method with mapped hour and month.

Methods (With Hour and Months)	RMSE with Standard Deviation
XGBoost	480.530 ± 7.650
Random Forest	487.682 ± 13.435
Linear Regression	774.602 ± 6.826

Discussion & Conclusion

After including mapped hour and month, all the methods performance has been improved and gave lower RMSE value with Standard Deviation for all the methods. XGBoost gave the lower RMSE value than other approaches. So, for this research we concluded that XGBoost performs better than other machine learning approaches.

References

[1] Akuzmiakova A.K., Colasg G.C. and McKeehan A. "Short-term Memory Solar Energy Forecasting at University of Illinois", Stanford University, 2017.
[2] Yinghao Chu, Bryan Urquhart, Seyyed M.I., Gohari, Hugo, T.C.Pedro, Jan Kleissl, Carlos and Coimbra F.M. "Short-term reforecasting of power output from a 48 MWe solar PV plant", Solar Energy, vol. 112, 2015, pp. 68-77.
[3]https://github.com/ColasGael/Machine-Learning-for-Solar-Energy-Prediction/tree/master/Datasets