WJERT, 2025, Vol. 11 Issue 11, 262-280. **Original Article** **ISSN 2454-695X**

World Journal of Engineering Research and Technology

www.wjert.org

Impact Factor: 7.029 **Coden USA: WJERA4**

EXTRATREES AND RANDOM FOREST MODELS FEATURE SELECTION PREDICTION

[1]*Isaac Raphael Okafor, [2]Garfield Jones, [3]Guangming Chen

[1*,2,3]Department of Industrial and Systems Engineering, Morgan State University, 1700 East Cold spring Lane, Baltimore, Maryland, USA.

Article Received on 07/10/2025 **Article Revised on 27/10/2025** **Article Published on 01/11/2025**

***Corresponding Author**
Isaac Raphael Okafor
Department of Industrial and Systems Engineering, Morgan State University, 1700 East Cold spring Lane, Baltimore, Maryland, USA.
https://doi.org/10.5281/zenodo.17541987

How to cite this Article: Isaac Raphael Okafor, Garfield Jones, Guangming Chen. (2025). Extratrees And Random Forest Models Feature Selection Prediction. World Journal of Engineering Research and Technology, 11(11), 262–280.

ABSTRACT

The importance of feature selection cannot be underrated due to its significance in dsetermining performance and reliability of predictive models. Machine learning techniques such as extratrees (ET) and random forest (RF) models were developed to leverage feature importance selection in predicting daily solar radiation (DSR) with precision. Daily solar radiation data for Southeastern (SE) Nigeria, such as temperature, relative humidity, precipitation, and wind speed were collected from the National Aeronautics and Space Administration (NASA) Prediction of Worldwide Energy Resource (POWER) database over a 10-year period ranging from 2012 to 2021 for the purpose of model development and testing. The findings reveal high impact of temperature and rank most important feature for the prediction of DSR by ET and RF models respectively in the five states. Among other features by the models, the temperature results are given as Abia- ET: Temp 0.275 and RF: Temp 0.267; Anambra- ET: Temp 0.278 and RF: Temp 0.293; Ebonyi- ET: Temp 0.285 and RF: Temp 0.266; Enugu- ET: Temp 0.274 and RF: Temp 0.267; Imo- ET: Temp 0.277 and RF: Temp 0.276. These results demonstrate great correlations in temperature values between the ET and RF models in the prediction of feature importance.

KEYWORDS: Feature Selection, Correlation, ET and RF Models, NASA POWER; DSR; Prediction.

NOMENCLATURES

CAMS = Copernicus Atmosphere Monitoring Service

DSR = Daily (Diurnal) Solar Radiation

ET = Extratrees Regressor

Extratrees = Extreme Randomized Trees

FEM = Finite Element Method

MSE = Mean Squared Error

ML = Machine Learning

NASA = National Aeronautics and Space Administration

Nig = Nigeria

NSRDB = National Solar Radiation Database

PPT= Precipitation

POWER = Prediction of Worldwide Energy Resource

PS = Surface Pressure

SE = Southeastern

SR = Solar Radiation

RF = Random Forest

RH = Relative Humidity

RK-KNN = Random Kernel- K- Nearest Neighbors

T = Temperature

TETFUND = Tertiary Education Trust Fund

WS = Wind Speed

W/m^2 = Watt per square meter.

1. INTRODUCTION

It will be incomplete studying solar radiation without considering the variables that influence its prediction. Considering all variables in predictive models introduces the challenge of the curse of dimensionality. Dimensionality refers to situation where models are negatively affected by the increase in the number of covariates. Some of these covariates may be giving redundant information which does not have any influence on prediction processes. These excess variables must be excluded when training the data. These call for variable selection in the order of importance. The question becomes, which variables are important to consider when developing a model? It is necessary to find ways of identifying those insignificant features and exclude them in model development before training the model. In addition, a high dimensional training data set can negatively affect a predictive model in several ways: (a) prediction accuracy is reduced, (b) models do not learn well a large number of irrelevant variables, (c) some important variables may not be picked due to interference from irrelevant variables, (d) makes the model complex to interpret; (e) the algorithm processing time is increased, (f) too many resources are used in the prediction process, (g) maintenance is difficult. According to Hossain et al. (2013), including an optimal feature subset provides better prediction accuracy in predicting solar power, and the selection of a minimal feature set giving the best possible classification results is desirable for practical reasons. The subset fits well the data because it contains the most important variables. An optimal subset reduces overfitting in the model development process. Different variable selection methods have been used to find this optimal subset of features, but the methods were developed to suit different data conditions.

As the primary renewable energy source, a thorough understanding of the various aspects of global solar radiation is critical and extremely beneficial for various applications, including architectural design, agricultural meteorology, weather prediction, climate monitoring, health, and even tourism applications and research (Oliver and Jackson, 2001; Jiang et al., 2019; Chen et al., 2012, Fu and Rich, 2002). Although solar radiation components are one of the most frequently measured meteorological variables, the number of measuring sites remains limited, especially in emerging and underdeveloped countries. Additionally, certain measurements may be incorrect due to equipment maintenance and calibration issues (Kaba, Sangul and Kandimaz, 2018). Nevertheless, some excellent alternatives and projects have provided various solar radiation and other meteorological data for various regions worldwide, such as the National Solar Radiation DataBase (NSRDB) (Yang, 2021; Zhang et al., 2019; Gueyamard, Habte and Sengupta, 2018), the Prediction Of Worldwide Energy Resources (POWER) (Mensour et al., 2017), the Copernicus Atmosphere Monitoring Service (CAMS) (Benanrou, Quardouz, Allaouzi and Ahmed, 2020; Ettayyebi and Himdi, 2018), and the global meteorological database METEONORM (Jallah et al., 2020; Bendali et al., 2020).

RF model was developed to improve learning performance by use of a voting system which enables them to measure the importance of variables as well as predict the target/response variable. RF classifies features as important or rejected. Leo Breiman (2001) asserted that RFs are the best classifiers for high dimensional data. They form an ensemble of weak unbiased classifiers which combine their results during final classification. No tuning is necessary since trees are grown until each leaf contains just a few elements. RF entails the creation of an ensemble comprising numerous decision trees to generate a unified and more precise prediction or outcome (Niklas, 2024).

Munshi and Moharil (2022) added that RFs can handle missing values with no overfitting. They are less affected by noise in the data, robust to outliers, and stable. Villegas-Mier et al. (2022) discovered that RFs gave a robust performance with similar results in two different scenarios. RFs delivered accurate and precise results when mapping solar radiation (SR) at high latitudes (Babar et al., 2020). Lee et al. (2020) also found that lagged SR features contribute significantly to the ensemble model. Their RF model produced SR at one-hour lag, relative humidity and showed to have high importance scores on USA data from six stations. Zeng et al. (2020) concluded that the RF model had high performance under different climates and geographic conditions. The importance scores computed by Ibrahim and Khatib (2017) showed that sunshine, hour and temperature were the most important features. We appreciate that assessing importance scores guides the feature selection process. However, the concept of ignoring correlations makes RFs insensitive to interaction effects.

The ExtraTrees (ET) model, also referred to as Extremely Randomized Trees, represents a machine learning ensemble technique that aggregates forecasts generated by numerous independent decision trees. This approach serves to amplify variance, mitigate overfitting, and takes into account crucial hyperparameters such as n estimators, max features, and min samples split.

ET constitutes a sophisticated machine learning algorithm that is predicated upon the foundational concepts of decision trees and ensemble learning, similar to Random Forests. Its architecture is meticulously engineered to enhance predictive accuracy and computational efficiency through the incorporation of increased randomness within the tree-construction mechanism. This methodology proves especially advantageous in the management of intricate datasets and the augmentation of model generalization. (Shu et al., 2022).

According to Shu et al (2022), the ET machine learning algorithm is employed in the research article to forecast the asymmetric warpage geometry associated with panel-level packaging. By constructing a training database utilizing the finite element method (FEM) for various geometric dimensions of fan-out panel-level packaging models, ET can effectively predict warpage values across a spectrum of package sizes. The algorithm plays a significant role in mitigating the issues related to asymmetric warpage, which may be induced by elements such as irregular epoxy shrinkage. Through the application of ensemble learning, ET provides a rapid and precise technique for estimating warpage in panel-level packaging, thereby facilitating the enhancement of packaging design and manufacturing methodologies.

According to Shu et al (2022), some of the key features of ET include the following:

- Augmented Randomness: In contrast to conventional decision trees or even RF, ET incorporate randomness not solely in the selection of data subsets but also in the determination of split points for each feature. This enhanced randomness contributes to a reduction in variance and overfitting, thereby rendering the model more resilient to data noise.
- Efficiency in Managing Extensive Datasets: The algorithm exhibits computational efficiency, as it eschews the necessity for bootstrapping the data, a procedure commonly employed in Random Forests. Rather, it utilizes the complete dataset to construct each tree, which may result in expedited training times and superior utilization of the available data.

ETs have been effectively implemented in various contexts, such as forecasting asymmetric warpage in panel-level packaging. This underscores the algorithm's proficiency in addressing intricate, real-world challenges by capitalizing on its ensemble learning methodology to enhance predictive accuracy.

Following Hossain et al. (2013) and for the present study, a comparative investigation of temperature, wind speed, relative humidity and precipitation is carried out using feature importance analysis to achieve the objective. A comparison of the variable selection methods is necessary before prediction involving training and testing with the models.

The main contribution of the present study is to demonstrate to the solar energy industry, meteorologists and the body of knowledge at large that feature importance selection approach perform differently on different environmental data sets; for instance, the case of the five southeastern states Nigeria.

2. MATERIALS AND METHODS

2.1 Data Collection and Analysis

Meteorological data such as daily air temperature (Temp), relative humidity (RH), precipitation (PPT), and wind speed (WS) spanning a 10-year period from 2012 to 2021 were obtained from the National Aeronautics and Space Administration (NASA) and was used to investigate the contributions of each of the variables to produce solar radiation. To prevent issues like overfitting, underfitting, and bias, the collected data underwent training and testing through K-fold cross-validation. The depiction of the various input data employed in forecasting daily solar radiation in the Southeastern region of Nigeria is illustrated in Figure 1.

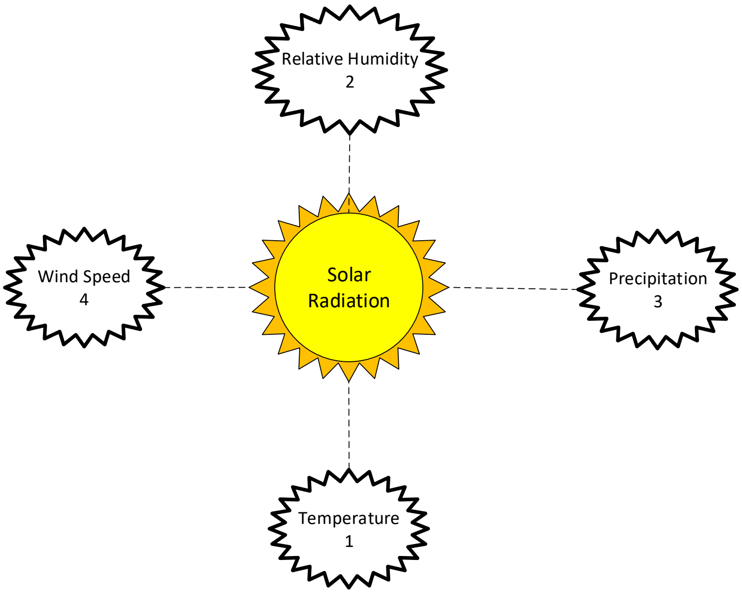


Figure 1: Input variables for daily solar radiation.

2.2 Latitude and Longitude of the Study Area

Latitude affects the angle of solar incidence, which in turn influences the amount of solar radiation received. For example, locations closer to the equator receive more direct sunlight, resulting in higher solar radiation levels compared to those at higher latitudes (Chabane et al., 2020) (Shruthi et al., 2017). Longitude impacts the timing of solar radiation due to the Earth's rotation, affecting the duration of sunlight exposure throughout the day (Chabane et al., 2020).

Table 1: Latitude and Longitude of the Study Area.

State	Elevation (m)	Latitude (Deg.)	Longitude (Deg.)
Abia	92.75	5.4566	7.5306
Anambra	103.94	6.2309	6.9337
Ebonyi	88.44	6.2664	8.0145
Enugu	151.33	6.4463	7.4893
Imo	62.54	5.5675	7.0519

2.3 Feature Visualization

Feature visualization helps to visually generate a summarized version of the data to demonstrate the relationships between data parameters, to study the changes of the different factors, and/or to demonstrate how these properties affect the variable under research (Brush, 2020). This feature data selection visualization can be seen using Matplotlib package in Google Colab 3 Jupyter Notebook. These are shown below:

2.3.1 Feature Importance Scores and Ranking

Feature importance is a measure of how much a feature contributes to the prediction made by a machine learning model. In simpler terms, it tells you which features are the most influential in determining the outcome. Feature importance helps us identify the most impactful features, leading to more efficient, interpretable, and high-performing models. According to Brownlee (2016), feature importance refers to strategies that assign a value to input parameters on the basis of their predictive power for a target variable.

Feature importance scores are critical components of a predictive modelling project because they provide insights into the data, insights into the model, and the foundation for dimensionality reduction and feature selection, which can improve the efficiency and effectiveness of a predictive model on the problem (Brownlee, 2016). Such scores are valuable and may be used on various situations in a predictive modeling issue, including the following: 1. understanding the data completely. 2. improving the comprehension of a model. 3. reducing the number of input features. In this paper, feature importance algorithms values such as RF and ET were obtained from Google Colab using Jupyter Notebook Python 3 and are shown in the table 1 below.

Table 1 shows comparison of mean square error (MSE) for RF and ET models to validate the feature importance performance obtained from the Python 3 Google Colab.

Table 2: Comparison of Feature Importance based on Performance Evaluation Metric for the Random Forest and Extratrees Models.

Location	Model	Metric	Feature Importance				
		MSE	Temp	WS	RH	PPT	Feature Ranking
Abia	ET	**3.65**	**0.2746**	0.2689	0.2403	0.2165	Temp
	RF	3.64	**0.2674**	0.2562	0.2542	0.2222	Temp
Anambra	ET	**3.93**	**0.2778**	0.2506	0.2438	0.2280	Temp
	RF	3.88	**0.2928**	0.2377	0.2400	0.2296	Temp
Ebonyi	ET	**4.04**	**0.2850**	0.2394	0.2627	0.2130	Temp
	RF	3.98	0.2663	0.2358	**0.2987**	0.1992	RH
Enugu	ET	**4.14**	**0.2740**	0.2654	0.2509	0.2097	Temp
	RF	4.02	**0.2668**	0.2628	0.2624	0.2080	Temp
Imo	ET	**3.67**	**0.2774**	0.2672	0.2436	0.2118	Temp
	RF	3.63	**0.2759**	0.2515	0.2477	0.2249	Temp

Mean Square Error (MSE) is a common performance metric that impacts feature importance in prediction models by permuting (randomly shuffling) the values of a feature and observing the change in the model's prediction error. A higher MSE after permuting a feature value indicates a greater importance of that feature. As can be seen in table 1; for example, temperature as a feature is the most ranked feature as indicated in the temperature values and the MSE values for the RF and ET models respectively.

2.3.2 Architecture for Feature Importance

Figure 2 shows the factors that enhance the efficacy of solar radiation prediction. Two main steps are involved in the architecture. First, the feature importance of the datasets is computed with RF and ET models. Second, the effect of the features- Temp, RH, WS and PPT is investigated. Here, these features are integrated into the models in order to obtain the highest accuracy in ranking. Computation of the datasets helps to select the best/ top ranked features and train the models. The training set was 90% and 10% testing.

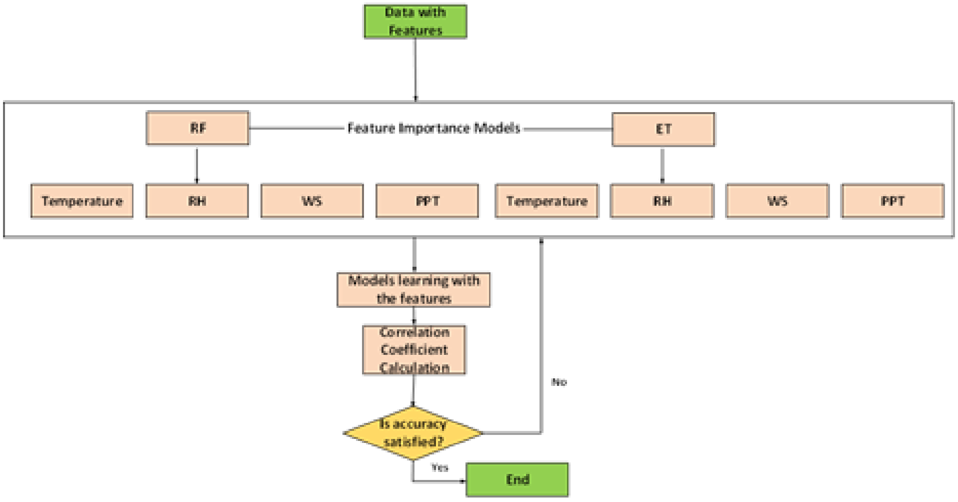


Figure 2: Architecture for Feature Importance.

2.3.3 Correlation Heat Map Feature Importance

A heat map is a data visualization technique that represents values using different colors, typically in a gradient from cool to warm tones. The intensity of the color corresponds to the magnitude of the value, making it easier to identify patterns, correlations, and hotspots within the dataset. The heat map feature reveals the most influential factors in the prediction daily solar radiation for the southeast Nigeria. The correlation heat map feature importance analysis can be found in figures 14-18.

2.4 K-Fold Cross Validation

Cross-validation is a statistical method primarily used in machine learning to estimate the skill of a machine learning model on unseen data. It is commonly used in applied machine learning to compare and select a model for a given predictive modeling problem because it is easy to understand, easy to implement, and results in skill estimates that generally have a lower bias than other methods. K-fold cross-validation is a technique for evaluating predictive models. However, the value of k depends on the size of the dataset. A good standard values for k in k-fold cross-validation are 5 and 10. However, the value of k depends on the size of the dataset. For example, for small datasets, we can use higher values for k. Larger values of k will increase the runtime of the cross-validation algorithm and the computational cost. When a specific value for k is chosen, it may be used in place of k in the reference to the model, such as k=10 becoming 10-fold cross-validation: 10% of the test set is held back each time. When k=5, it implies that 20% of the test set is held back each time and 80% for data training. The general procedure for K-fold cross-validation when k=10 is shown in figure 3 below.

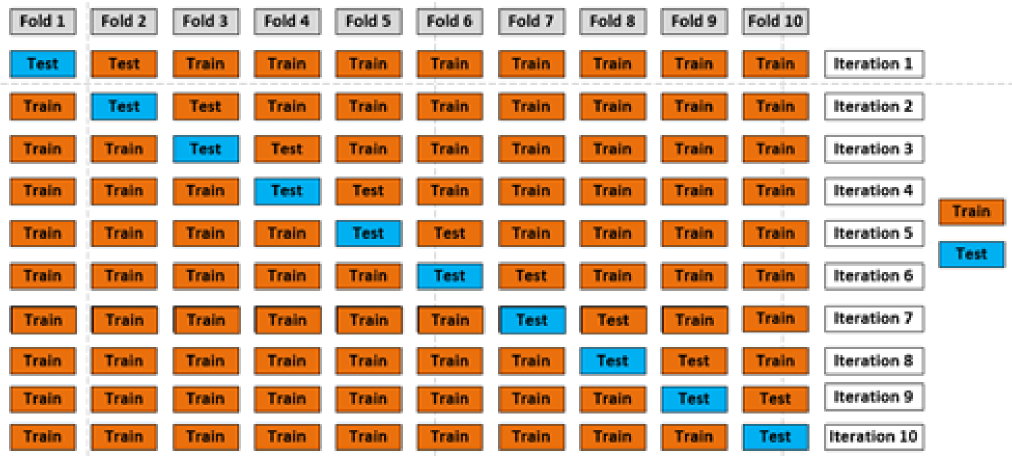


Figure 3: The general procedure for K-fold cross-validation when K=10.

From figure 3, k=10, it means that dataset is divided into 10 folds. One-fold run is the train and the other run is the test. 90% training and 10% testing of the datasets.

3. RESULTS AND DISCUSSION

3.1 Random Forest and ExtraTrees Feature Importance Development

Table 3 shows the first four values of the dataset

Table 3: The header of the dataset.

TEMP	RH	PPT	WS	SR
24.12	75.56	0.23	2.94	11.77
22.97	65.81	0.00	2.73	11.89
22.76	68.44	0.00	2.51	10.94
22.14	65.25	0.02	3.41	10.85
22.72	70.19	0.18	1.79	10.23

Table 4 shows the last four values of the dataset

Table 4: The End of the dataset.

TEMP	RH	PPT	WS	SR
26.45	83.88	0.10	1.51	10.68
24.47	78.50	0.71	1.62	10.17
23.95	72.50	0.31	2.34	8.53
24.54	72.62	0.00	2.44	12.62
24.76	73.56	3.09	1.80	11.17

3.2 Feature Conditions/Correlation for Increased Solar Radiation for each of the States

Table 5 shows feature conditions/correlation for increased solar radiation for each of the States obtained from correlation heat map in figs 14-18.

Table 5: Feature Conditions/Correlation for Increased Solar Radiation for each of the States.

State	Condition				
Abia	Temp(0.25)	WS(-0.29)	PPT(-0.21)	RH(-0.15)	Increased SR
Anambra	Temp(0.25)	WS(-0.23)	PPT(-0.12)	RH(-0.08)	Increased SR
Ebonyi	Temp(0.20)	WS(-0.18)	PPT(-0.23)	RH(-0.20)	Increased SR
Enugu	Temp(0.18)	WS(-0.21)	PPT(-0.17)	RH(-0.13)	Increased SR
Imo	Temp(0.26)	WS(-0.30)	PPT(-0.19)	RH(-0.16)	Increased SR

From figures 4-8 and figs 9-13, the bar charts show that with random forest and extratrees model feature importance, temperature and wind speed are the most important weather parameters that impact the daily solar radiation in four states of the southeast Nigeria for the period 2012-2021, except for Ebonyi state where relative humidity (RH) and temperature (Temp) are the most critical features.

Figures 14-18 show correlation heat map analysis where temperature (Temp) and wind speed (WS) are the most important weather parameters that impact the daily solar radiation in four states of the southeast Nigeria for the period 2012-2021, except for Ebonyi state where relative temperature (Temp) and precipitation (PPT) are the most critical features.

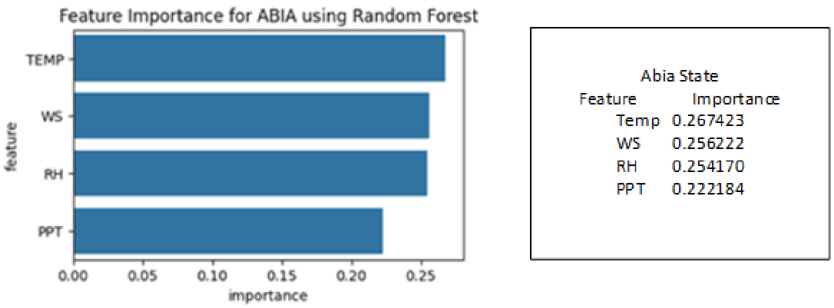


Figure 4: Feature Importance for Abia State using Random Forest Model.

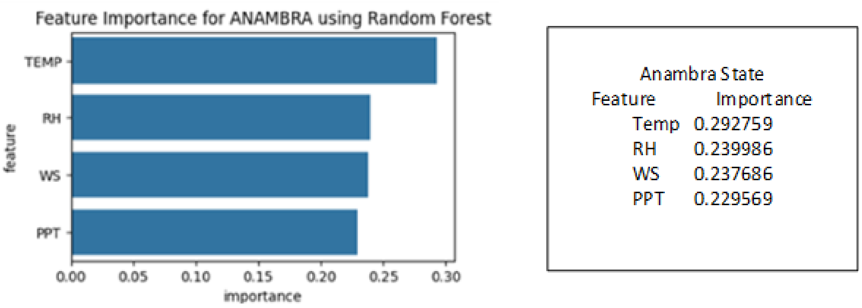
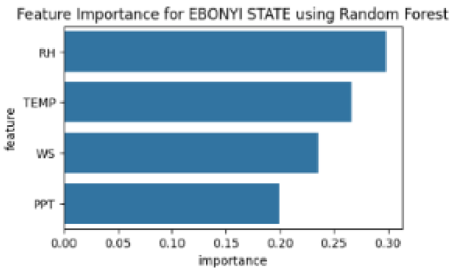
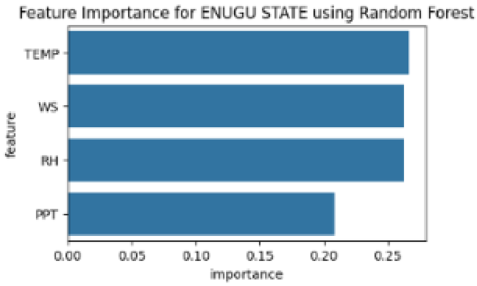


Figure 5: Feature Importance for Anambra State using Random Forest Model.



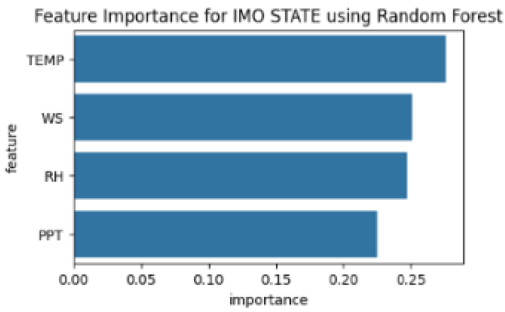
Ebonyi State	
Feature	Importance
RH	0.298677
Temp	0.266306
WS	0.235834
PPT	0.199183

Figure 6: Feature Importance for Ebonyi State using Random Forest Model.



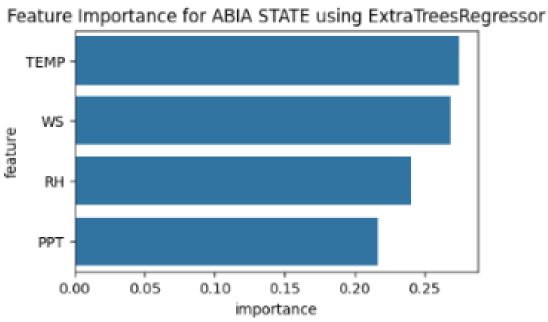
Enugu State	
Feature	Importance
Temp	0.266828
WS	0.262798
RH	0.262370
PPT	0.208004

Figure 7: Feature Importance for Enugu State using Random Forest Model.



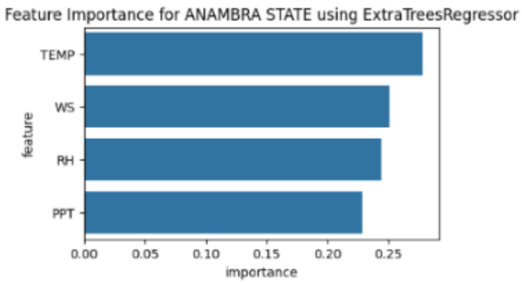
Imo State	
Feature	Importance
Temp	0.275855
WS	0.251464
RH	0.247746
PPT	0.224936

Figure 8: Feature Importance for Imo State using Random Forest Model.



Abia State	
Feature	Importance
TEMP	0.274643
WS	0.268576
RH	0.240263
PPT	0.216517

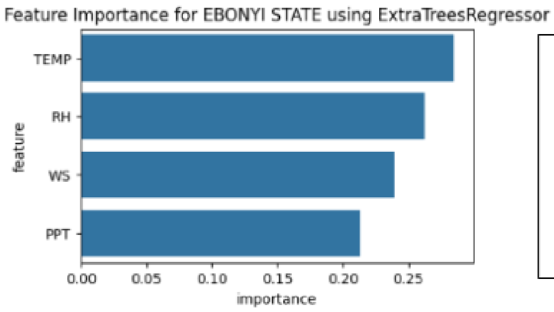
Figure 9: Feature Importance for Abia State using Extratrees Model.



Anambra State

Feature	Importance
TEMP	0.277772
WS	0.250553
RH	0.243753
PPT	0.227922

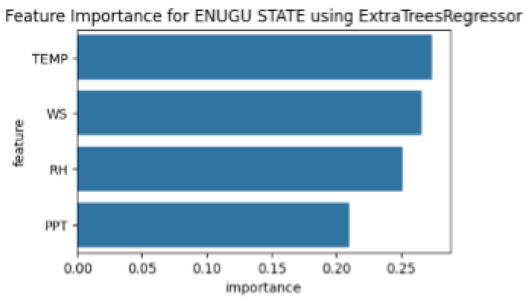
Figure 10: Feature Importance for Anambra State using Extratrees Model.



Ebonyi State

Feature	Importance
TEMP	0.284979
RH	0.262689
WS	0.239385
PPT	0.212947

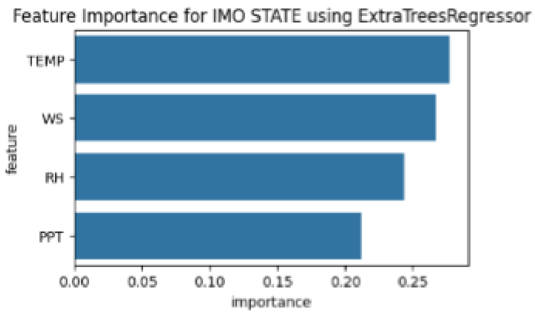
Figure 11: Feature Importance for Ebonyi State using Extratrees Model.



Enugu State

Feature	Importance
TEMP	0.274017
WS	0.265423
RH	0.250888
PPT	0.209672

Figure 12: Feature Importance for Enugu State using Extratrees Model.



Imo State

Feature	Importance
TEMP	0.277413
WS	0.267170
RH	0.243584
PPT	0.211833

Figure 13: Feature Importance for Imo State using Extratrees Model.

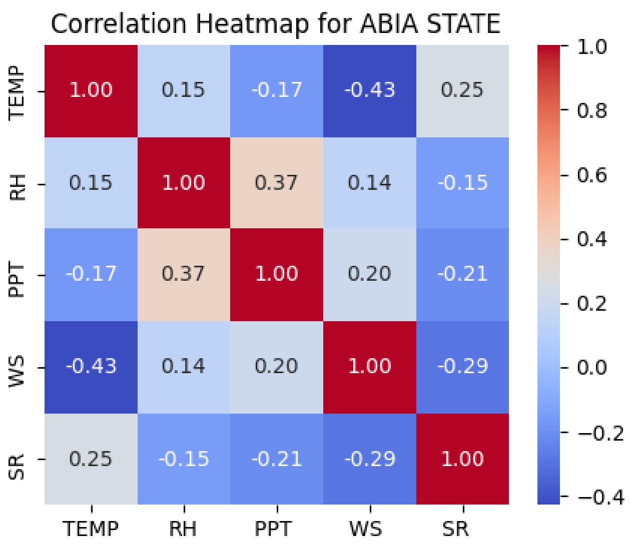


Figure 14: Correlation Heat Map Feature for Abia State.

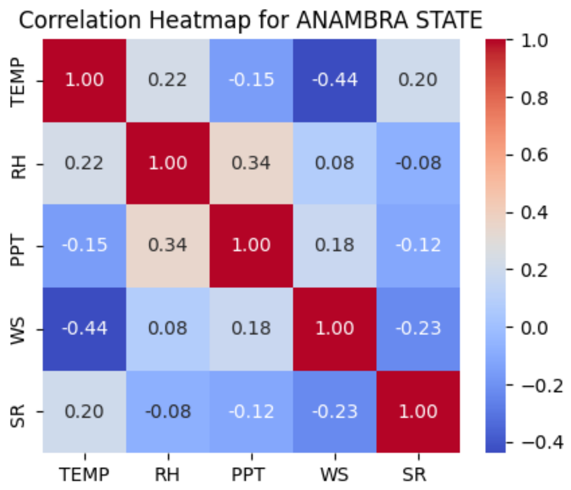


Figure 15: Correlation Heat Map Feature for Anambra State.

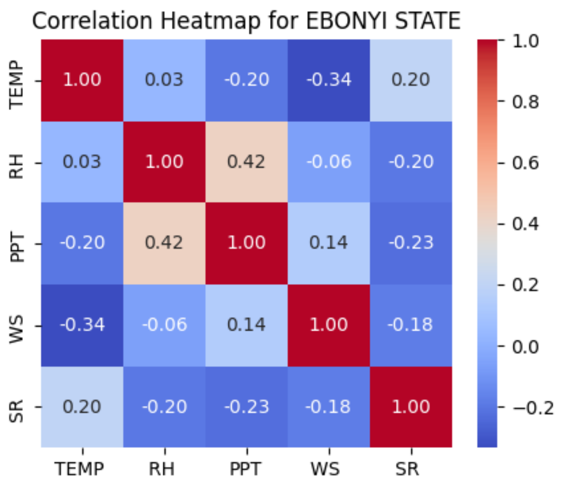


Figure 16: Correlation Heat Map Feature for Ebonyi State.

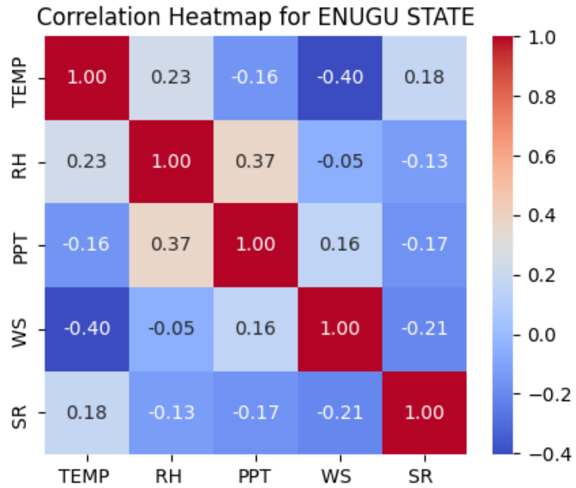


Figure 17: Correlation Heat Map Feature for Enugu State.

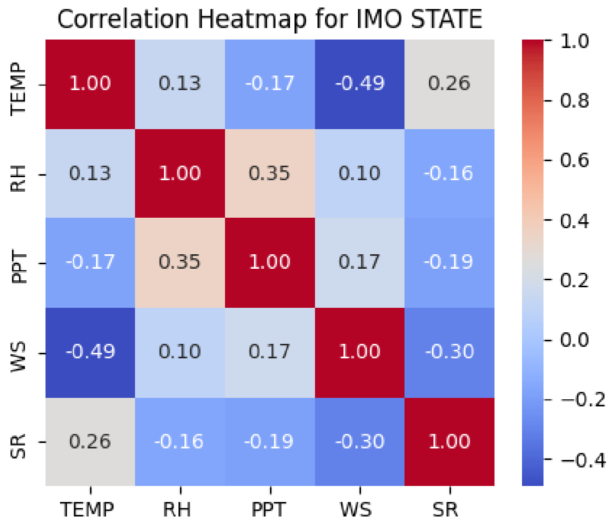


Figure 18: Correlation Heat Map Feature for Imo State.

CONCLUSION

Feature importance defines which features are the most influential in predicting solar radiation in the region of study. In other words, it helps to identify the most impactful features, leading to more efficient, interpretable, and high-performing models.

With the results obtained in the heat map analysis in figures 14-18, temperature, relative humidity, precipitation, and wind speed can be said to be strongly correlated with solar radiation because they contribute to increased solar radiation. As shown in table 2, temperature is ranked the most impactful feature to the prediction of solar radiation in the region with RF and ET models. ET and RF are both ensemble learning methods that rely on decision trees, but have some key differences when it comes to feature importance prediction.

Table 2 also shows that RF model is more accurate, stable and reliable for feature importance ranking as indicated in the MSE values; For example, RF MSE values is a little lower than the ET MSE values. ET model is better for computational efficiency and fast training time when dealing with very large datasets. Table 5 shows the conditions for increased solar radiation for each of the states obtained from figures 14-18.

With random forest and extratrees model feature importance, temperature and wind speed are the most important weather parameters that impact the daily solar radiation in four states of the southeast Nigeria for the period 2012-2021, except for Ebonyi state where relative humidity (RH) and temperature (Temp) are the most critical features.

ACKNOWLEDGEMENTS

Thank you to the Tertiary Education Trust Fund (TETFUND), Nigeria, for funding my doctoral program in the Department of Industrial and Systems Engineering at Morgan State University, Baltimore, Maryland, USA, and to NASA's Prediction of Worldwide Energy Resource (POWER) for data collection from their Database-Data Access Viewer. I am grateful to my advisor, Dr. Jones Garfield, who taught me IEGR 501-Introduction to Advanced Systems Engineering in the fall semester of 2022 and made me accept this task owing to his interest in engineering optimization. All my dissertation committee members and my academic mentor, Dr. Guangming Chen, have inspired my Ph.D. in Industrial and Systems Engineering at Morgan State University. I also want to thank Dr. Anita Pandey from Morgan State University for her online class, "Final Project for All." Her class has helped me grasp dissertation proposal writing.

REFERENCES

1. Babar, B., Luppino, L.T., Bostrom, T. and Anfinsen, S.N. Random forest regression for improved mapping of solar irradiance at high latitudes. Solar Energy, 2020; 198: 81-92.
2. Benamrou, B., Ouardouz, M., Allaouzi, I., and Ahmed, M. B. A proposed model to forecast hourly global solar irradiation based on satellite derived data, deep learning and machine learning approaches, J. Ecol. Eng., 2020; 21, 4: 26– 28.
3. Bendali, W., Saber, I., Bourachdi, B., Boussetta, M., and Mourad, Y. Deep learning using genetic algorithm optimization for short term solar irradiance forecasting, in Proc. 4th Int. Conf. on Intelligent Computing in Data Sciences (ICDS), Fez, Morocco, 2020; 1–8.
4. Breiman, L. Random Forests Machine Learning, 2001; 45: 5–32. https://doi.org/10.1023/A:1010933404324

5. Brownlee, J. Feature importance and feature selection with XGBoost in python, Machine Learning Mastery, 2016. https://machinelearningmastery.com/feature-importance and-feature-selection-with-xgboost-in-python/.
6. Brush, K. Data visualization, tech target search business analytics, 2020. https://searchbusinessanalytics.techtarget.com/ definition/data-visualization.
7. Chabane, F., Arif, A., Benramache, S. Prediction of the Solar Radiation Map on Algeria by Latitude and Longitude Coordinates. Tecnica Italiana-Italian Journal of Engineering Science, 2020; 64: 213-215. 10.18280/ti-ijes.642-413.
8. Chen, S., Kharrazi, A., Liang, S., Fath, B, D., Lenzen, M., Jinyue Yan, J. Advanced approaches and applications of energy footprints toward the promotion of global sustainability, Applied Energy, 2020; 261: 114415, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2019.114415.
9. Ettayyebi, H and El Himdi, K. Artificial neural networks for forecasting the 24 hours ahead of global solar irradiance, AIP Conf. Proc., 2018; 2056(1): 020010.
10. Fu, P., and Rich, P.M. A geometric solar radiation model with applications in agriculture and forestry, Computers and Electronics in Agriculture, 2002; 37, 1–3: 25-35, ISSN 0168-1699, https://doi.org/10.1016/S0168-1699(02)00115-1. (https://www.sciencedirect.com/science/article/pii/S0168169902001151)
11. Gueymard, C. A., Habte, A., and Sengupta, M. Reducing uncertainties in large-scale solar resource data: The impact of aerosols, IEEE J. Photovolt., 2018; 8(6): 1732–1737.
12. Hossain, M. R., Than, A. M. and Shawkat Ali, A. B. M. The Effectiveness of feature selection method in solar power prediction. Journal of Renewable Energy: 2013. http://dx.doi.org/10.1155/2013/952613.
13. Jallal, M. A., El Yassini, A., Chabaa, S., Zeroual, A., and Ibnyaich, S. A deep learning algorithm for solar radiation time series forecasting: A case study of El Kelaa des Sraghna city, Rev. d'Intell. Artif., 2020; 34(5): 563–569.
14. Jiang, Liping & Wang, Caixia & Xie, Guohui & Li, Zhu & Zhang, Hongpeng & Pei, Zheyi & Zeng, Pingliang & Kang, Chongqing & Zhang, Ning & Li, Qionghui & Hu, Anping. Variable-Generation Integration in China: An Update. IEEE Power and Energy Magazine, 2019; 17: 99-107. 10.1109/MPE.2019.2932640.
15. Kaba, K., Sarıgül, M., Avcı, M. H., Kandırmaz, M. Estimation of daily globa solar radiation using deep learning model, Energy, 2018; 162: 126-135, ISSN 0360-5442, https://doi.org/10.1016/j.energy.2018.07.202. (https://www.sciencedirect.com/science/article/pii/S0360544218315007)

16. Khatib, T. and Ibrahim, I.A. A novel hybrid model for hourly global solar radiation prediction using random forests technique and firefly algorithm. Energy Conversion and Management, 2017; 138: 413-425.
17. Lee, j., Wang, W., Harrou, F. and Sun, Y. Reliable solar irradiance prediction using ensemble learning-based models: A comparative study. Energy conversion and management, 2020; 208: 112-582.
18. Mensour, O. N., Bouaddi, S., Abnay, B., Hlimi, B., and Ihlal, A. Mapping and estimation of monthly global solar irradiation in different zones in Souss-Massa area, Morocco, using artificial neural networks, Int.J. Photoenergy, 2017; 8547437.
19. Munshi, A. and Moharil, R.M. Solar radiation forecasting using random forest. AIP Conference Proceedings, 2022; 2424: 050003. DOI.org/10.1063/5.0076827.
20. Niklas, D. Random Forest: A Complete Guide for Machine Learning. All you need to know about the random forest model in machine learning, 2024.
21. Oliver, M. and Jackson, T. The market for solar photovoltaics, Energy Policy, 1999; 27(7): 371-385, ISSN 0301-4215, https://doi.org/10.1016/S0301-4215(99)000385. (https://www.sciencedirect.com/science/article/pii/S0301421599000385)
22. Shruti V. P., Padhye, S.M., Padhye, A.D. Environmental determinants of species turnover of uatic Heteroptera in freshwater ecosystems of the Western Ghats, India, Limnologica, 2020; 80: 125730, ISSN 0075-9511, https://doi.org/10.1016/j.limno.2019.125730. (https://www.sciencedirect.com/science/article/pii/S0075951119300842)
23. Shu, Z., Wang, B., Kou-Ning, Chiang. Using Extra Trees Machine Learning Algorithm to Predict the Asymmetric Warpage Geometry of Panel Level Packaging, 2022. doi: 10.1109/IMPACT56280.2022.9966670.
24. Villegas-Mier, C. G., Rodriguez-Resendiz, J., Alvarez-Alvarado, J.M., Jimenez-Hernandez, H. and Odry, A. Optimized Random Forest for Solar Radiation Prediction Using Sunshine Hours. Micromachines, 2022; 13: 1406.
25. Yang, D. Z. Validation of the 5-min irradiance from the national solar radiation database (NSRDB), J. Renew. Sustain. Energy, 2021; 13(1): 016101.
26. Zeng, Z., Wang, Z., Gui, K., Yan, X., Gao, M., Luo, M., Geng, H., Liao, T., Li, X., An, J., Liu, H., He, C., Ning, G. and Yang, Y. Daily global solar radiation in China estimated from high density meteorological observations: A random forest model framework. Earth and Space Science, 2020; 7: e2019EA001058. DOI. org/10.1029/2019EA001058.

27. Zhang, W., Kleiber, W., Florita, A.R., Hodge, B.M, Barry Mather. Modeling and Simulation of High-Frequency Solar Irradiance. IEEE Journal of Photovoltaics, 2019; 9: 1.