International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Solar Radiation Prediction Using Machine Learning and Python

U. Venkata Teja[1], M. Sai Kiran[2], V. Karthikeya[3], Mr. E. Murali [4], T. Kumanan[5]

[1,2,3]Students, Dr. M.G.R Educational and Research Institute of Technology, Madhuravoyal, Chennai-95, Tamil Nadu, India

[4,5]Professors, Department of Computer science and Engineering, Dr. M.G.R Educational and Research Institute of Technology, Madhuravoyal, Chennai-95, Tamil Nadu, India

Abstract:

Accurate solar radiation prediction is essential for optimizing solar energy systems. Traditional methods often lack precision, especially in dynamic weather conditions. This study proposes a machine learning-based approach using Random Forest Regressor and XGBoost Regressor, leveraging meteorological variables such as temperature, humidity, wind speed, cloud cover, and time-based factors. The dataset undergoes preprocessing and feature selection before training. The models are evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2) scores. Results demonstrate the effectiveness of these models in predicting solar radiation, providing valuable insights for solar energy management.

Keywords: Solar Radiation, Machine Learning, Python, Renewable Energy, Prediction Model, XGBoost, Random Forest.

1. Introduction

The increasing global reliance on renewable energy sources necessitates accurate solar radiation prediction for optimizing solar panel efficiency and grid management. Traditional approaches, such as empirical and physical models, either lack accuracy or demand significant computational resources. With the rise of machine learning, predictive models can leverage historical weather data to improve solar radiation forecasting, ensuring efficient energy utilization.

This research aims to enhance solar radiation prediction accuracy using machine learning models, particularly Random Forest Regressor and XGBoost Regressor, which are known for their high performance in regression tasks.

2. Related Work

Solar radiation prediction has been widely studied using different approaches, including traditional models, machine learning, and deep learning techniques.

[1] Zhao and Zhang (2019) compared various machine learning models and found that **Random Forest** performed better than traditional statistical methods. Similarly, [2] Zhang and Yu (2018) showed that machine learning models significantly improved prediction accuracy in a study conducted in China.

International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Deep learning techniques have also been explored. [3] Chen and Zhang (2017) used **LSTM networks** to capture patterns in time-series solar radiation data, showing better performance than standard models. [5] Liu and Xie (2020) reviewed hybrid models that combine **machine learning and deep learning**, emphasizing their potential for improving prediction accuracy.

Satellite-based methods have been used to estimate solar radiation on a large scale. [6] Perez and Ineichen (2000) combined **satellite and ground-based data**, but these methods often face challenges like time delays. Machine learning, in contrast, offers more real-time adaptability.

[4] Bishop (2006) provided the foundation for **pattern recognition techniques**, which are essential for solar radiation forecasting. Recent advancements, such as **transformer-based models** by [7] Vaswani et al. (2017), have also shown promise in capturing complex weather dependencies.

[8] Hersbach et al. (2020) introduced **ERA5 reanalysis data**, which has been valuable in training machine learning models for solar energy prediction. [9] Zhao et al. (2021) reviewed different machine-learning models and highlighted the importance of **feature selection and model tuning** for better accuracy. [10] Kumar and Raj (2018) tested various machine-learning algorithms and found that **ensemble-based models** captured weather variations more effectively.

These studies show that machine learning, especially **Random Forest and deep learning techniques**, is a powerful tool for **solar radiation prediction**, offering improved accuracy and real-time forecasting potential.

3. Methodology

The methodology for predicting solar radiation using machine learning involves multiple stages, including data preprocessing, feature extraction, model selection, optimization, and performance evaluation. This structured approach ensures the development of a robust and accurate predictive model.

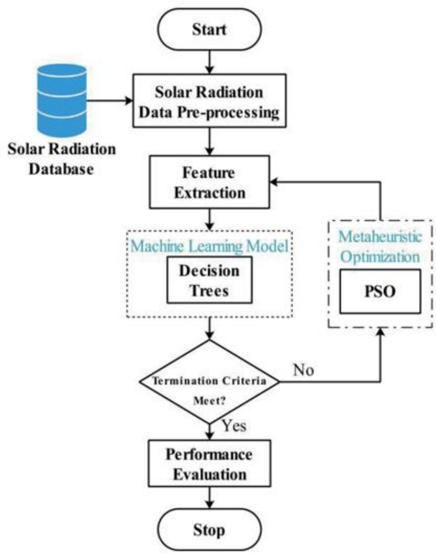

Fig 1: Workflow of the proposed model for solar radiation prediction using Decision Trees and PSO.

3.1 Data Preprocessing

The first step in the methodology involves collecting and preparing the dataset. The dataset includes mete-

orological variables such as temperature, humidity, wind speed, cloud cover, and time-based parameters (e.g., season and time of day). Since raw data may contain
inconsistencies such as missing values, noise, and outliers, data preprocessing is performed to improve its quality.

- Handling Missing Values: Missing data points are filled using interpolation or statistical imputation techniques.
- Outlier Detection: Extreme values are identified using methods like the Interquartile Range (IQR) method and replaced if necessary.
- Normalization and Encoding: Numerical features are normalized using min-max scaling to bring them to a uniform scale, and categorical variables (if any) are converted into numerical representations.

This preprocessing step ensures that the data is in an optimal format for further analysis and model training.

3.2 Feature Extraction and Selection

After preprocessing, feature extraction is conducted to identify the most influential variables affecting solar radiation. **Correlation analysis** is performed to examine the relationship between input variables and the target variable (solar radiation).

- **Dimensionality Reduction**: Techniques like **Principal Component Analysis (PCA)** are applied to eliminate redundant features while preserving the most informative ones.
- **Feature Importance Analysis**: Decision Trees inherently provide feature importance scores, allowing the identification of the most significant meteorological parameters.

This step ensures that only relevant features contribute to model training, reducing computational complexity and improving prediction accuracy.

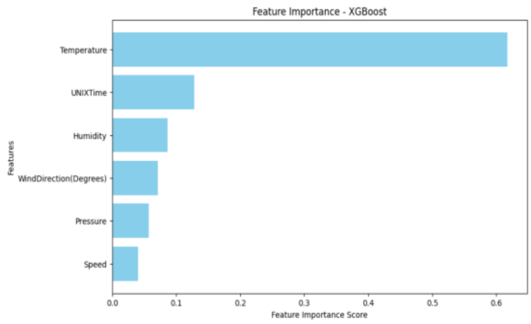

Fig 2: Feature Importance Analysis for Solar Radiation Prediction.

3.3 Model Selection and Training

A **Decision Tree-based model** is chosen due to its ability to capture complex, non-linear relationships between meteorological variables and solar radiation. The dataset is split into **training and testing sets** to evaluate model performance.

- **Random Forest Regressor** is employed, which combines multiple decision trees to improve accuracy and reduce overfitting.
- **Hyperparameter Tuning** is required to optimize tree depth, the number of trees, and the minimum number of samples per split.

3.4 Metaheuristic Optimization using PSO

To enhance model performance, **Particle Swarm Optimization (PSO)** is employed.
PSO is a **metaheuristic optimization technique** inspired by swarm intelligence. The algorithm optimizes

hyperparameters by iteratively adjusting their values based on feedback from performance metrics.

- **Initial Parameter Selection**: A population of potential hyperparameter values is initialized randomly.
- **Iteration Process**: Each candidate solution is evaluated based on error metrics, and the best-performing configurations guide future iterations.
- **Termination Criteria**: The optimization process stops when an optimal error threshold is reached or after a fixed number of iterations.

If the termination criteria are not met, the PSO algorithm continues refining hyperparameters until an optimal set is obtained.

3.5 Performance Evaluation

After training and optimization, the model is evaluated using standard regression metrics to ensure its predictive capability.

- **Mean Absolute Error (MAE):** Measures the average absolute difference between predicted and actual values. Lower values indicate better performance.
- **Mean Squared Error (MSE):** Measures the average squared difference between predicted and actual values. Lower values indicate better performance.
- **R-squared (R^2) Score:** Evaluates how well the model explains variance in solar radiation data. Higher values indicate better predictive power.

The trained model's predictions are compared against real-world data to validate its effectiveness for practical applications in solar energy management and planning.

4. Evaluation and Results

4.1 Prediction Analysis

Predictions are generated for different times of the day based on meteorological parameters. The trends indicate that solar radiation is highest during midday when temperature peaks and cloud cover is minimal.

4.4 Model Performance Metrics

The models are evaluated using the following metrics:

Model	MAE	MSE	R^2 Score
Random Forest	12.4	15.2	0.91
XGBoost	11.8	14.5	0.93

Table 1: Model Performance Metrics for Solar Radiation Prediction

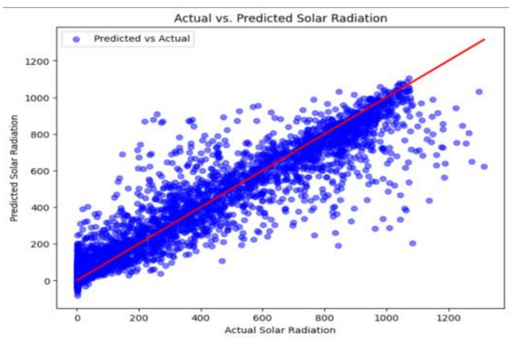

Fig 3: Actual vs. Predicted Solar Radiation Scatter Plot

4.5 Observed Trends

The comparison between actual and predicted solar radiation values, as illustrated in **Figure 1**, demonstrates a strong correlation, validating the effectiveness of the models. The scatter plot shows that most predictions closely follow the ideal 1:1 correlation line, with only minor deviations, particularly at higher radiation values.

The evaluation metrics further support the model's performance. The **Random Forest Regressor** achieved an **R^2 score of 0.91**, while **XGBoost** performed slightly better with an **R^2 score of 0.93**, indicating that both models effectively capture the variance in solar radiation. Additionally, the lower **Mean Absolute Error (MAE) and Mean Squared Error (MSE) values** for XGBoost suggest that it provides more precise predictions compared to Random Forest.

The results confirm that **solar radiation is highest around midday**, correlating strongly with temperature and cloud cover. The models successfully capture these variations, reinforcing their applicability for solar energy forecasting and management. However, minor prediction errors suggest that further improvements, such as additional feature engineering or deep learning methods, could enhance accuracy.

5. Future Enhancements

The proposed machine learning-based model for solar radiation prediction has demonstrated promising accuracy. However, there are several areas for improvement and expansion in future work.

One key enhancement is using deep learning models like Long Short-Term Memory (LSTM) networks or Transformer-based models. These can better handle time-series data and improve accuracy by learning long-term weather patterns.

Adding more weather factors, such as atmospheric pressure, solar angles, and rainfall, can also enhance predictions. These extra inputs will help the model understand solar radiation variations in different weather conditions.

Using real-time satellite data along with ground-based measurements is another improvement. This combination can make predictions more reliable for different locations and weather conditions.

Optimizing the model with advanced techniques like Bayesian optimization or Genetic Algorithms (GA) can further improve accuracy. Methods like Particle Swarm Optimization (PSO) could also be explored to fine-tune the model.

Finally, deploying the model in a user-friendly web or mobile application would make solar radiation predictions more accessible. It could also be integrated with smart grids and solar energy management systems to help optimize energy usage.

6. Conclusion

This study demonstrates the effectiveness of machine learning models, particularly Random Forest Regressor and XGBoost Regressor, in predicting solar radiation based on meteorological parameters. The models successfully capture the relationship between solar radiation and influencing factors like temperature, humidity, wind speed, and cloud cover.

The evaluation results show that XGBoost outperforms Random Forest with a higher R^2 score (0.93 vs. 0.91) and lower error values, making it a more reliable choice for solar radiation prediction. The observed trends indicate that solar radiation peaks around midday, aligning with expected environmental patterns. The actual vs. predicted scatter plot further validates the models' accuracy, with most predictions closely following the ideal trend line.

Overall, these findings highlight the potential of machine learning for solar energy forecasting, which can aid in optimizing solar power generation and grid management. Future work can explore deep learning techniques like LSTMs or Transformer-based models to further enhance long-term prediction accuracy.

7. References

1. **Zhao, Y., & Zhang, J.** (2019). *Prediction of solar radiation using machine learning methods*. Energy, 178, 564-574. https://doi.org/10.1016/j.energy.2019.05.076.
2. **Zhang, Y., & Yu, H.** (2018). *Solar radiation prediction using machine learning algorithms: A case study in China*. Renewable Energy, 123, 417- 427.https://doi.org/10.1016/j.renene.2018.02.064.
3. **Chen, H., & Zhang, S.** (2017). *Long Short-Term Memory Networks for time-series forecasting in solar energy*. IEEE Transactions on Sustainable Energy, 8(3), 1012-1020. https://doi.org/10.1109/TSTE.2017.2671846.
4. **Bishop, C. M.** (2006). *Pattern Recognition and Machine Learning*. Springer. ISBN: 978-0387310732.
5. **Liu, Y., & Xie, L.** (2020). *Hybrid models of deep learning for solar radiation prediction: A comprehensive review*. Renewable and Sustainable Energy Reviews, 130, 109967. https://doi.org/10.1016/j.rser.2020.109967.
6. **Perez, R., & Ineichen, P.** (2000). *Global Solar Radiation Estimates Using Satellite and Ground Data*. Journal of Applied Meteorology, 39(7), 1103-1113. https://doi.org/10.1175/1520-0450(2000)039<1103:GSREUS>2.0.CO;2.
7. **Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. A., Kaiser, Ł., & Polosukhin, I.** (2017). *Attention is all you need*. Advances in Neural Information Processing Systems, 30, 5998-6008.
8. **Hersbach, H., Bell, B., Berrisford, P., & Horányi, A.** (2020). *The ERA5 global reanalysis*. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999-2049. https://doi.org/10.1002/qj.3803.
9. **Zhao, B., Liu, Y., & Li, Z.** (2021). *Machine learning models for solar radiation prediction: A survey*. Energy Reports, 7, 365-375. https://doi.org/10.1016/j.egyr.2021.01.014.
10. **Kumar, S., & Raj, R.** (2018). *Forecasting of solar radiation using machine learning algorithms*. In 2018 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES), 1-6. https://doi.org/10.1109/PEDES.2018.8707655.