Solar Energy Site Suitability Mapping Using Satellite Data and Machine Learning

WiDS Final Project Report

Name: Prince Bhati and Harshita Joshi
Institution: Indian Institute of Technology Bombay
Mentor: Tanvi Anand
Project Duration: 6 Weeks
Project Type: Guided WiDS Project

1. Introduction

The increasing global demand for clean and sustainable energy has made renewable energy sources such as solar power critically important. India, particularly states like Rajasthan, receives high solar irradiance throughout the year, making it ideal for large-scale solar energy deployment. However, selecting optimal locations for solar power plants requires careful evaluation of multiple geospatial factors such as terrain, land cover, and environmental constraints.

Traditional site selection methods rely heavily on manual surveys and limited datasets, which are not scalable for large regions. Recent advances in satellite remote sensing and machine learning (ML) provide an efficient alternative by enabling automated, data-driven site suitability analysis.

This project develops a **solar energy site suitability mapping framework** using satellite-derived geospatial data and machine learning techniques. Google Earth Engine (GEE) is used for large-scale geospatial data processing, while a Random Forest classifier is employed to identify suitable and non-suitable locations for solar energy installations across Rajasthan, India

2. Study Area

The study area selected for this project is **Rajasthan, India**, one of the most solar-rich states in the country. Rajasthan receives high solar radiation for most of the year and already hosts major solar parks such as Bhadla and Jodhpur, making it suitable for supervised learning approaches where known solar plant locations can be used as reference data.

The state exhibits diverse geographical and land-use characteristics, including desert regions, agricultural land, urban settlements, and forested areas. This diversity allows the machine learning model to learn patterns associated with both suitable and unsuitable locations for solar energy deployment.

The analysis is conducted at the **state level**, enabling large-scale assessment while maintaining meaningful spatial resolution for site suitability mapping

3. Literature Review and Site Selection Parameters

Previous studies on solar energy site suitability emphasize the importance of integrating multiple geospatial parameters rather than relying on a single factor. Literature reviewed during the project suggests that a **multi-criteria approach** significantly improves site suitability assessment accuracy.

Key parameters considered in this study include:

- **Solar Radiation:** Determines energy generation potential
- **Slope and Elevation:** Flat or gently sloping terrain reduces construction and maintenance costs
- **Land Cover:** Helps exclude forests, water bodies, and dense urban areas
- **Environmental Constraints:** Protected and agricultural lands are typically unsuitable

These parameters form the foundation for feature selection and model development in this project

4. Datasets Used

All datasets were accessed and processed using **Google Earth Engine (GEE)**.

Parameter	Dataset	Source
Solar Radiation	ERA5 / MODIS	Google Earth Engine
Elevation	SRTM DEM	GEE
Slope	Derived from DEM	GEE
Land Cover	ESA WorldCover	GEE
Existing Solar Sites	Open datasets & literature	Research papers
Wind Speed (10 m)	ERA5 (u & v components)	GEE

Due to accessibility constraints for some solar radiation variables, **wind speed at 10 m height** was additionally explored using ERA5 monthly data. Wind speed was derived from u and v wind components and visualized over Rajasthan as a supplementary renewable energy indicator

5. Methodology

5.1 Overall Workflow

1. Define study region (Rajasthan) in Google Earth Engine
2. Import relevant satellite datasets
3. Preprocess and derive geospatial features
4. Label known solar plant locations
5. Train a supervised ML model
6. Validate model performance
7. Generate final solar suitability map

5.2 Feature Engineering

Satellite datasets were processed using GEE reducers to extract statistical summaries such as mean values. Terrain attributes like slope were derived from elevation data. Land cover datasets were used to mask unsuitable regions such as water bodies and dense forests.

For wind energy analysis, wind speed was computed using the formula:

Wind Speed=*u*2+*v*2Wind Speed=u2+v2

where *u* and *v* represent wind components at 10 m height.

Extracted features were exported in tabular format for machine learning model training

5.3 Machine Learning Model

A **Random Forest classifier** was used due to its robustness, ability to handle non-linear relationships, and effectiveness with high-dimensional geospatial data. The model classifies regions into **suitable** and **non-suitable** categories for solar energy installation.

Model performance was evaluated using:

- Accuracy
- Precision
- Recall

- Confusion Matrix

6. Results and Analysis

The trained Random Forest model successfully identified regions with high suitability for solar energy deployment. Areas with flat terrain, favorable land cover, and high renewable energy potential were classified as suitable. The generated suitability map aligns well with existing large-scale solar installations in Rajasthan, validating the effectiveness of the approach.

Wind speed visualization further highlighted regions with additional renewable energy potential, demonstrating the extensibility of the framework to multi-energy analysis.

7. Challenges Faced

- Learning Google Earth Engine and geospatial scripting
- Feature engineering for satellite-based datasets
- Managing project timelines alongside academic commitments

These challenges were addressed through continuous learning and iterative experimentation

8. Future Scope

- Integration of higher-resolution solar radiation datasets
- Hyperparameter tuning and comparison with other ML models
- Multi-criteria decision analysis (MCDA) integration
- Expansion to pan-India or global-scale mapping

9. Conclusion

This project demonstrates the effectiveness of combining satellite remote sensing data with machine learning for large-scale solar energy site suitability mapping. The proposed framework provides a scalable and data-driven approach for identifying optimal renewable energy sites and can support sustainable energy planning and policy decisions.