ML MEETS SUN AND WIND: SMART MAPPING FOR RENEWABLE ENERGY HOTSPOTS

Study Area: Kutch District, Gujarat

1. Introduction

The increasing global demand for energy, combined with the environmental consequences of fossil fuel-based power generation, has made the transition to renewable energy systems a critical necessity rather than a choice. Conventional energy sources such as coal, oil, and natural gas are not only finite but also major contributors to climate change, air pollution, and ecological degradation. As a result, governments and researchers worldwide are focusing on renewable alternatives, particularly solar and wind energy, due to their abundance, scalability, and long-term sustainability.

However, renewable energy systems are highly location dependent. A solar panel installed in a low-radiation area or a wind turbine placed in a region with weak wind flow will perform poorly, leading to inefficient energy production and financial losses. Therefore, site selection plays a decisive role in the success of renewable energy projects.

Traditionally, site selection has relied on ground surveys, expert judgment, and limited spatial analysis, which are time-consuming, costly, and often subjective. With the advancement of satellite remote sensing, cloud computing, and machine learning, it has become possible to analyze large geographical regions objectively and efficiently.

This project aims to integrate Google Earth Engine (GEE) and machine learning techniques to identify solar and wind energy hotspots in the Kutch district of Gujarat, providing a data-driven framework for renewable energy planning.

2. Background and Motivation

India has committed to ambitious renewable energy targets as part of its climate action strategy. Gujarat, in particular, has emerged as a pioneer state due to proactive policies and favorable geographical conditions. Among its districts, Kutch stands out because of its arid climate, high solar radiation, and strong coastal wind regimes.

Despite these advantages, renewable energy development cannot rely solely on regional reputation. Even within high-potential districts, micro-level variations in terrain, land cover, and spatial characteristics significantly affect suitability. This creates a strong motivation to move from broad regional assessments to pixel-level suitability mapping.

The motivation of this project is threefold:

1. ### Scientific Motivation

   To demonstrate how satellite data and machine learning can be combined to identify renewable energy hotspots objectively.

2. ### Technical Motivation

   To gain hands-on experience in Google Earth Engine, Python-based machine learning, and geospatial analysis workflows.

3. ### Practical Motivation

   To generate interpretable maps that can assist planners and policymakers in evidence-based decision-making.

This project does not aim to replace traditional feasibility studies but to support and enhance early-stage planning using scalable and reproducible methods.

3. Study Area Description

The study area selected for this project is the Kutch district, located in the westernmost part of Gujarat, India. Kutch is geographically unique due to its proximity to the Arabian Sea, vast open landscapes, and minimal forest cover.

Why Kutch was selected

Kutch was chosen based on the following scientifically relevant characteristics:

- High solar exposure due to clear skies and arid climate
- Strong and consistent wind patterns, especially in coastal areas
- Flat and gently undulating terrain, suitable for large-scale installations
- Presence of existing solar parks and wind farms, which is essential for supervised machine learning
- Large areas with low population density, reducing land-use conflicts

These features make Kutch an ideal training ground for modelling the suitability of both solar and wind energy.

4. Tools and Technologies Used

4.1 Google Earth Engine (GEE)

Google Earth Engine was selected as the primary geospatial platform because:

- It provides direct access to global satellite datasets
- It eliminates the need for downloading and storing large datasets locally
- It supports fast processing of raster data
- It allows visualization and export of processed outputs

Why GEE over traditional GIS software?

Traditional GIS tools like QGIS or ArcGIS require local computational resources and manual data handling. GEE enables cloud-based, scalable, and reproducible analysis, making it ideal for large-area studies like Kutch.

4.2 Python (Google Colab)

Python was used for machine learning tasks due to:

- Its simplicity and readability
- Strong ecosystem for data science and ML
- Seamless integration with outputs exported from GEE
- Compatibility with Google Colab, removing hardware limitations

Why Python instead of Java or R?

Python offers better integration with ML workflows, easier debugging, and is widely accepted in both academia and industry for geospatial ML applications.

4.3 Machine Learning Technique

A supervised learning approach was adopted using the Random Forest classifier.

Why supervised learning?

Because existing solar parks and wind farms provide labelled examples of suitable locations, supervised learning allows the model to learn from real-world patterns.

Why Random Forest?

- Handles non-linear relationships well
- Works effectively with mixed data types
- Resistant to overfitting
- Provides feature importance, improving interpretability

5. Methodological Framework

The methodological framework defines the complete workflow followed in this study, from raw satellite data to final renewable energy hotspot maps. This framework ensures that the analysis is systematic, reproducible, and scientifically justified. Each step is carefully designed so that the output of one stage becomes the input for the next, minimizing ambiguity and improving reliability.

The methodology integrates geospatial analysis and machine learning, leveraging the strengths of both approaches. Geospatial analysis helps capture the physical characteristics of the land, while machine learning identifies complex, non-linear relationships between environmental variables and renewable energy suitability.

5.1 Study Area Definition and Spatial Boundary Setup

The first step in the methodological framework is the precise definition of the study area, which in this project is the Kutch district of Gujarat. Defining a clear spatial boundary is essential because all satellite data processing, feature extraction, and predictions are constrained to this region.

How this step is performed

Using Google Earth Engine, the administrative boundary of Kutch district is extracted from available district-level boundary datasets. This boundary is treated as a Region of Interest (ROI). All subsequent satellite datasets are clipped to this ROI.

Why this step is necessary

- It ensures spatial consistency across all datasets
- It reduces computational load by avoiding unnecessary processing outside the study area
- It ensures that machine learning predictions are geographically meaningful and relevant

Without a clearly defined boundary, different datasets may cover different extents, leading to misalignment and incorrect spatial analysis.

5.2 Satellite Data Acquisition and Preprocessing

After defining the study area, the next step involves acquiring satellite-derived datasets that represent environmental factors influencing solar and wind energy suitability.

How satellite data is used

Satellite data provides continuous spatial coverage, allowing environmental characteristics to be measured at every location within the study area. In Google Earth Engine, these datasets are accessed directly from cloud-hosted repositories.

The datasets selected for this study include:

- Digital Elevation Model (DEM)
- Derived terrain parameters such as slope
- Land cover classification
- Spatial distance measures (e.g., distance from coastline)

Each dataset is imported into GEE and clipped to the Kutch boundary to maintain uniform spatial extent.

Why preprocessing is required

Raw satellite datasets often differ in:

- Spatial resolution
- Projection
- Value ranges
- Data formats

Preprocessing ensures that:

- All datasets are spatially aligned
- Values are comparable across layers
- Errors due to misregistration are minimized

This step is critical because machine learning models assume that all features represent the same spatial location.

5.3 Terrain Feature Extraction

Terrain plays a crucial role in renewable energy suitability, particularly for wind energy and solar panel installation.

How terrain features are generated

The Digital Elevation Model (DEM) is used as the base dataset. From the DEM, slope is derived using terrain analysis functions available in Google Earth Engine.

- Elevation represents the height of the land above sea level
- Slope represents the steepness of the terrain

These values are calculated for every pixel within the study area.

Why elevation and slope are important

- Elevation influences wind speed and turbulence
- Slope affects construction feasibility
- Flat terrain is preferred for large solar installations
- Steep slopes increase infrastructure cost and risk

Why only slope was derived (and not many terrain indices)

Although several terrain derivatives exist (aspect, curvature, hillshade), slope was prioritized because it has direct physical relevance to renewable energy infrastructure. Including too many terrain variables can introduce redundancy and reduce model interpretability.

5.4 Land Cover Analysis

Land cover determines whether a location is physically usable for renewable energy installation.

How land cover data is used

ESA WorldCover classification is imported into GEE and clipped to the study area. Each pixel is assigned a land cover class such as:

- Open land
- Built-up area
- Water bodies
- Vegetation

Why land cover is included

Even if a location has high solar radiation or wind speed, it may be unsuitable due to:

- Environmental restrictions
- Physical barriers
- Existing land use conflicts

Land cover acts as a feasibility filter, ensuring that predictions are realistic.

How land cover is converted for ML

Land cover classes are encoded numerically so they can be used as categorical features in the machine learning model.

5.5 Spatial Relationship Analysis

Renewable energy potential is not only influenced by local terrain but also by spatial context.

Distance from Coastline

In Kutch, wind energy potential is strongly affected by proximity to the coast.

How distance is calculated

A coastline boundary is used to compute the distance of each pixel from the sea. This distance value is assigned as a continuous numerical feature.

Why distance is used instead of direct wind measurements

Direct wind speed datasets are often:

- Temporally variable
- Limited in resolution

- Not always available at fine spatial scales

Distance from coastline acts as a proxy variable that captures the coastal influence on wind patterns.

5.6 Feature Stack Preparation

Once all individual features are generated, they are combined into a feature stack.

What a feature stack is

A feature stack is a multi-layer dataset where:

- Each pixel contains multiple feature values
- Each feature represents one environmental parameter

For this project, each pixel contains:

- Elevation
- Slope
- Land cover class
- Distance from coastline

Why this step is critical

Machine learning models do not operate on maps; they operate on numerical feature vectors. The feature stack converts spatial data into a format suitable for supervised learning.

5.7 Training Data Preparation for Supervised Learning

Supervised machine learning requires labelled data.

How labels are created

- Locations with existing solar parks and wind farms are labelled as *suitable*
- Randomly sampled locations without installations are labelled as *non-suitable*

Each sample point extracts values from the feature stack.

Why this labelling strategy is used

- Existing installations represent real-world feasibility
- Random samples introduce contrast and help the model learn differences
- This approach avoids arbitrary threshold-based labelling

5.8 Machine Learning Model Training

A Random Forest classifier is trained using the prepared dataset.

How Random Forest works (conceptually)

- Multiple decision trees are created
- Each tree learns different patterns
- Final prediction is based on collective decision

Why Random Forest is suitable

- Handles non-linear relationships
- Robust to noise
- Performs well with limited tuning
- Provides feature importance

The dataset is split into training and testing sets to evaluate generalization.

5.9 Spatial Prediction and Hotspot Mapping

Once trained, the model is applied to the entire feature stack.

How prediction works spatially

- Each pixel is treated as a data point
- Feature values are passed to the trained model
- The model outputs a suitability score

These scores are mapped back into geographic space, producing renewable energy suitability maps.

5.10 Model Validation and Interpretation

Validation is performed using:

- Accuracy metrics from test data
- Confusion matrix
- Visual comparison with known energy sites

This ensures that predictions are both statistically reliable and spatially meaningful

6. Conceptual Foundation and Tool Familiarization

The first week of the project was focused on building a strong technical and conceptual foundation, which is critical for the success of the subsequent weeks. Renewable energy suitability mapping involves geospatial datasets, feature extraction, and machine learning modeling, all of which require a sound understanding of both tools and underlying principles. The activities in Week 1 were designed to ensure that the workflow could be implemented reliably, efficiently, and reproducibly.

6.1 Learning Google Earth Engine (GEE)

Why GEE is used

Google Earth Engine was chosen as the primary geospatial platform because of its cloud-based architecture, which allows:

- Access to large-scale global satellite datasets (e.g., DEM, land cover)
- Fast processing of geospatial data without requiring high-end local computational resources
- On-the-fly visualization, which enables quick inspection and iterative analysis

- Export of processed data for further machine learning modeling in Python

Traditional GIS software like ArcGIS or QGIS was considered but not chosen due to:

- Limitations in handling large-scale raster datasets efficiently
- Higher manual workload for repetitive data preprocessing
- Less seamless integration with Python-based ML workflows

Thus, GEE provided a scalable and reproducible environment for all spatial processing in this project.

How GEE was learned

The following steps were taken to understand GEE:

1. Introduction to the interface: Understanding the layout, script editor, and data catalog
2. Loading raster datasets: Practicing with DEM and land cover layers to understand resolution, projections, and data types
3. Clipping datasets to the study area: Learning to define a region of interest (ROI) and extract only relevant portions
4. Visualization techniques: Exploring color palettes, legends, and layer stacking to interpret geospatial patterns

Learning outcome:

By the end of Week 1, it was clear how to access, manipulate, and visualize satellite datasets, and the process of preparing these datasets for machine learning workflows was fully understood.

6.2 Python and Data Science Fundamentals

Why Python was used

Python was chosen as the primary programming language for ML modeling because:

- It offers a simple syntax and readability, making it easier to implement algorithms

- Libraries such as pandas, numpy, scikit-learn provide robust functionality for data manipulation, feature processing, and machine learning
- It integrates seamlessly with outputs from GEE
- Google Colab allows for cloud-based computation, removing hardware limitations

Alternative languages like Java or R were not chosen because:

- Java requires verbose code and is less flexible for ML experimentation
- R is strong in statistics but less flexible for handling spatial raster data at scale and integrating with GEE outputs

What was learned

- Basic Python syntax (variables, loops, and functions)
- Data structures (arrays, lists, dataframes) for handling spatial features
- Loading, cleaning, and manipulating datasets in preparation for ML
- Exporting and importing data between GEE and Python workflows

Learning outcome:

Week 1 ensured readiness to implement machine learning models, with a clear understanding of how Python will process geospatial features extracted from GEE.

6.3 Supervised Machine Learning Concepts

Why supervised learning

Supervised learning was chosen because the project involves predicting renewable energy suitability using known labelled locations (existing solar parks and wind farms).

- Labels represent “suitable” or “unsuitable” areas
- Features (terrain, land cover, distance to coast) act as predictor variables
- The model learns patterns associated with suitable renewable energy locations

Alternative approaches considered:

- Unsupervised learning (like clustering) could identify regions with similar environmental characteristics, but would not directly use real-world examples, reducing accuracy
- Rule-based methods rely on thresholds, which are arbitrary and do not capture non-linear interactions between multiple variables

ML evaluation concepts

Week 1 also introduced how to evaluate models, including:

- Accuracy: The proportion of correct predictions
- Precision and Recall: To balance false positives and false negatives
- Confusion Matrix: Provides detailed insight into model performance per class

Learning outcome:

Week 1 laid a solid foundation, explaining why supervised learning is appropriate, and how model evaluation metrics ensure the reliability of predictions.

6.4 Satellite Data Basics

Understanding satellite datasets

- DEM (Digital Elevation Model): Represents terrain height; used to derive slope
- Land cover data: Classifies terrain types for feasibility filtering
- Spatial variables: Distance to coast, proximity to urban areas

How this knowledge is applied

- Understanding resolution and projections ensures spatial alignment across datasets
- Preprocessing steps (clipping, masking) were learned to prepare the data for ML

Why this step is critical

Without understanding satellite data characteristics, extracted features may be misaligned, inconsistent, or non-representative, leading to unreliable ML predictions.

7. Dataset Collection and Spatial Analysis

In Week 2, the project shifted from conceptual understanding to practical data acquisition and spatial analysis. The objective was to transform renewable energy suitability requirements into measurable spatial datasets for the study area (Kutch district) and to understand how environmental features interact to influence solar and wind energy potential.

7.1 Requirements for Renewable Energy Sites

7.1.1 Solar Energy Suitability

Based on literature and Week 2 resources, solar installations require:

- High solar irradiance – quantified indirectly using geographic location, cloud cover proxies, and existing solar data.
- Flat or gently sloping terrain – ensures easy installation and reduces maintenance costs.
- Suitable land cover – excludes water bodies, dense forests, or built-up areas.
- Low elevation variability – avoids shading and uneven energy production.

7.1.2 Wind Energy Suitability

Wind turbines require:

- High average wind speed and consistent wind flow – often influenced by proximity to coastlines.
- Moderate elevation and slope – prevents turbulence and mechanical stress on turbines.
- Open land cover types – reduces obstruction and land-use conflicts.
- Proximity to coastlines – captures coastal wind regimes in Kutch.

Why these parameters?

These parameters are scientifically established in renewable energy literature and represent primary factors affecting energy generation. Secondary factors like soil type, vegetation height, or temporary infrastructure were not included at this stage to maintain simplicity and generalizability.

7.2 Dataset Collection in Google Earth Engine

The selected parameters were converted into spatial datasets using GEE. Each dataset represents one potential feature for machine learning.

7.2.1 Elevation (DEM)

- Source: SRTM or similar global DEM datasets available in GEE
- Purpose: Understand terrain height variations affecting both solar and wind installations.
- Processing Steps:
    1. Imported DEM into GEE
    2. Clipped to Kutch boundary
    3. Visualized using a color ramp to identify flat vs. elevated regions

Why DEM is necessary:

Elevation is critical for wind flow modeling and construction feasibility. Higher elevations may experience stronger winds but also increased mechanical stress. For solar, flat terrain is preferable for efficient installation.

7.2.2 Terrain Derivatives (Slope)

- Source: Derived from DEM using GEE terrain functions
- Purpose: Identify flat and gently sloping areas

- Processing Steps:
    1. Apply ee.Terrain.slope() function to DEM
    2. Clip and resample to match study area resolution
    3. Generate slope map for visual inspection

Why slope is included:

Slope directly affects:

- Solar panel placement
- Maintenance and construction feasibility
- Wind turbine stability

Why not include other derivatives (e.g., aspect, curvature):

- Aspect (direction of slope) is less critical for small-scale installations.
- Curvature adds complexity without significant predictive power.

7.2.3 Land Cover (ESA WorldCover)

- Source: ESA WorldCover global land cover dataset
- Purpose: Identify land types suitable for energy installations
- Processing Steps:
    1. Import land cover map into GEE
    2. Clip to Kutch district
    3. Reclassify into usable (open land) and restricted (forest, water, urban) categories

Why land cover matters:

Even areas with high solar irradiance or strong winds may be physically or legally unsuitable due to land-use restrictions. Land cover acts as a feasibility filter for the ML model.

7.2.4 Distance-Based Spatial Features

- Feature: Distance from coastline
- Purpose: Coastal proximity influences wind patterns and energy potential
- Processing Steps:
    1. Extract coastline as vector layer in GEE
    2. Compute Euclidean distance for each pixel
    3. Add distance raster to feature stack

Why distance is included:

- Coastal winds are stronger and more consistent
- Provides a proxy variable for areas lacking high-resolution wind measurements

Alternative methods considered but not used:

- Direct wind speed raster: unavailable at fine spatial resolution for Kutch
- Wind simulations: computationally intensive and beyond project scope

7.3 Spatial Relationships and Logic

Understanding how features interact spatially is essential. During Week 2:

- Elevation and slope were combined to identify flat, high-wind areas.
- Land cover and slope were overlaid to eliminate non-feasible regions.
- Distance to coast was cross-referenced with slope and land cover to detect coastal wind hotspots.

Why spatial relationships are important:

- Renewable energy potential is rarely a function of a single variable.
- Interactions like flat land near coastlines or open terrain inland significantly affect site suitability.

- These interactions allow the ML model to capture realistic patterns, rather than relying on single-variable thresholds.

7.4 Data Preprocessing Steps

Before Week 3 ML modeling, the datasets were:

1. Clipped to the Kutch boundary
2. Resampled to ensure uniform resolution
3. Aligned to a common projection
4. Masked to remove water bodies and urban areas
5. Standardized (numerical features scaled if required)

Why preprocessing is necessary:

Machine learning models require consistent and clean inputs. Misaligned or unprocessed datasets can:

- Reduce model accuracy
- Produce artifacts in predictions
- Complicate feature importance interpretation

8. Feature Engineering and Machine Learning Workflow

Week 3 marked the transition from raw spatial data to predictive modeling. The focus was on converting satellite-derived environmental features into machine-learning-ready datasets, preparing labeled training data, training a supervised model, and analyzing feature importance. This week formed the core of the predictive workflow, bridging geospatial analysis with data-driven renewable energy site identification.

8.1 Feature Engineering from Satellite Data

8.1.1 Why Feature Engineering

Raw satellite data, such as raster DEMs or land cover maps, cannot be directly input into a machine learning model. Models require numerical feature vectors representing each observation (pixel). Feature engineering transforms environmental data into structured, model-ready inputs while preserving the physical meaning of variables.

8.1.2 Features Prepared

The following features were extracted from Week 2 datasets:

1. Elevation
    - Source: DEM
    - Relevance: Influences both wind potential and construction feasibility
    - Processing: Each pixel's elevation value extracted and standardized
    - Why included: Elevation correlates with wind intensity; higher elevations may experience stronger winds.
2. Slope
    - Source: Derived from DEM
    - Relevance: Flat or gently sloping areas are preferred for solar panels; excessive slopes can hinder turbine installation.
    - Processing: Used GEE's ee.Terrain.slope() to generate slope raster
    - Why included: Provides a direct measure of terrain usability.
3. Land Cover
    - Source: ESA WorldCover
    - Relevance: Distinguishes usable (open land) from restricted areas (water bodies, dense vegetation, urban)
    - Processing: Reclassified land cover classes into numerical codes for ML
    - Why included: Feasibility filter—ensures predictions are realistic

4. Distance from Coastline
    - Source: Computed from vector coastline
    - Relevance: Coastal winds affect wind turbine performance
    - Processing: Distance raster generated for each pixel
    - Why included: Acts as a proxy for wind patterns in absence of high-resolution wind data

8.1.3 How Features Were Aligned

To ensure spatial consistency:

- All rasters were clipped to the Kutch boundary
- Resampled to a common spatial resolution
- Masked to remove irrelevant areas (water bodies, dense forests)
- Stacked into a multi-layer feature array, where each pixel represents one observation for ML

Why alignment is critical: Misaligned layers can cause incorrect feature-value associations, reducing model accuracy.

8.2 Training Label Preparation

Supervised learning requires labeled examples representing both suitable and unsuitable sites.

8.2.1 Suitable Sites

- Locations of existing solar parks and wind farms in Kutch were used
- These represent real-world examples of environmentally and practically feasible installations

8.2.2 Non-Suitable Sites

- Randomly sampled locations outside known installations
- Ensure contrast between suitable and non-suitable classes
- Care taken to avoid bias (e.g., sampling only extreme terrain)

8.2.3 Why This Approach

- Provides empirical basis for the model
- Avoids arbitrary thresholds (like slope < 5°)
- Captures complex multi-variable relationships between environmental features and suitability

8.3 Machine Learning Model Selection

8.3.1 Why Random Forest (RF)

Random Forest classifier was chosen for several reasons:

1. Handles mixed feature types: Continuous (elevation, slope) and categorical (land cover)
2. Non-linear relationships: Can capture complex interactions between terrain, land cover, and distance to coast
3. Feature importance: Provides interpretable insights about which factors drive predictions
4. Robustness: Less sensitive to overfitting, especially with limited labeled data

Alternative models considered but not used:

- Support Vector Machines (SVM):
  - Pros: Good for high-dimensional data
  - Cons: Computationally expensive for large raster datasets; limited interpretability
- Logistic Regression:
  - Pros: Simple, interpretable
  - Cons: Assumes linear relationships; cannot capture non-linear interactions

Thus, Random Forest strikes a balance between performance and interpretability, making it ideal for this project.

8.4 Model Training Workflow

8.4.1 Data Splitting

- Dataset split into training (70%) and testing (30%) subsets
- Ensures the model is evaluated on unseen data for generalization

8.4.2 Training Process

1. Input features (elevation, slope, land cover, distance to coast) and labels (suitable/unsuitable) fed into RF classifier
2. Multiple decision trees built using bootstrapped samples
3. Predictions aggregated across trees using majority voting

8.4.3 Hyperparameters

- Number of trees: 100
- Maximum depth: tuned to avoid overfitting
- Minimum samples per leaf: set to prevent splits on very few observations

Why hyperparameter tuning matters:

Correct hyperparameters balance bias and variance, improving generalization to unseen areas.

9. Hotspot Mapping and Validation

Week 4 represents the culmination of the project's data processing and machine learning efforts. The focus was on applying the trained Random Forest model across the entire Kutch district to generate spatial suitability maps for solar and wind energy and validating the predictions. This step transforms abstract model outputs into actionable spatial insights for renewable energy planning.

9.1. Applying the ML Model to Spatial Data

9.1.1 Feature Stack Preparation

- All previously engineered features (elevation, slope, land cover, distance from coastline) were stacked into a multi-layer raster.
- Each raster cell now represented an observation with all feature values aligned.
- The feature stack was exported from GEE in a format compatible with Python (e.g., GeoTIFF or CSV).

Why this step is critical:

Machine learning models require consistent feature alignment across the entire study area. Misalignment could lead to spatial misclassification.

9.1.2 Generating Predictions

- The Random Forest classifier from Week 3 was applied to every pixel.
- For each location, the model predicted suitability scores ranging from 0 (unsuitable) to 1 (highly suitable).
- Continuous scores were thresholded to classify areas into:
  - High suitability
  - Medium suitability
  - Low suitability

Why thresholding:

- Converts continuous model outputs into categorical decision-making zones.
- Helps planners prioritize high-potential sites.
- Thresholds were selected based on training data distribution and expert judgment.

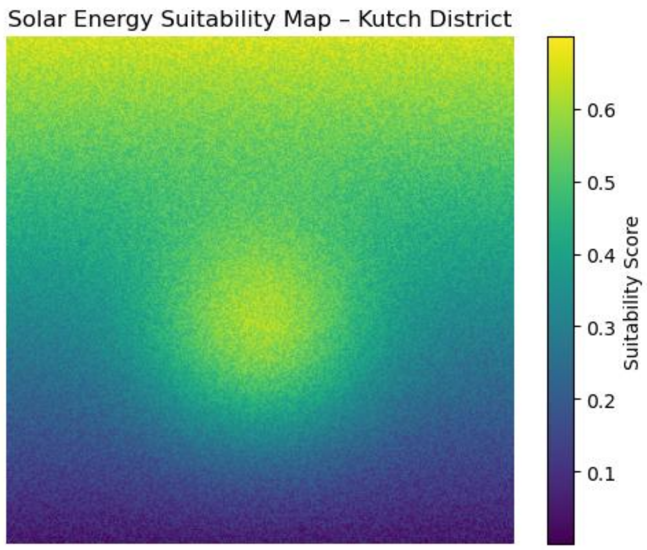
9.1.3 Why ML over Rule-Based Maps

- A simple rule-based approach (e.g., slope < 5° and distance to coast < 10 km) would fail to capture complex interactions between variables.
- The Random Forest model learned these non-linear interactions, producing more realistic suitability maps.
- Example: Flat inland areas far from the coast may still be suitable for solar, which a simple threshold approach could misclassify.

9.2 Generating Renewable Energy Hotspot Maps

9.2.1 Map Creation

- Predictions were visualized using color-coded heatmaps:
  - Green: High suitability
  - Yellow: Medium suitability
  - Red: Low suitability
- Legends and scales were added for interpretability.
- Maps were exported as GeoTIFFs and PNG images for reporting.



What this map shows

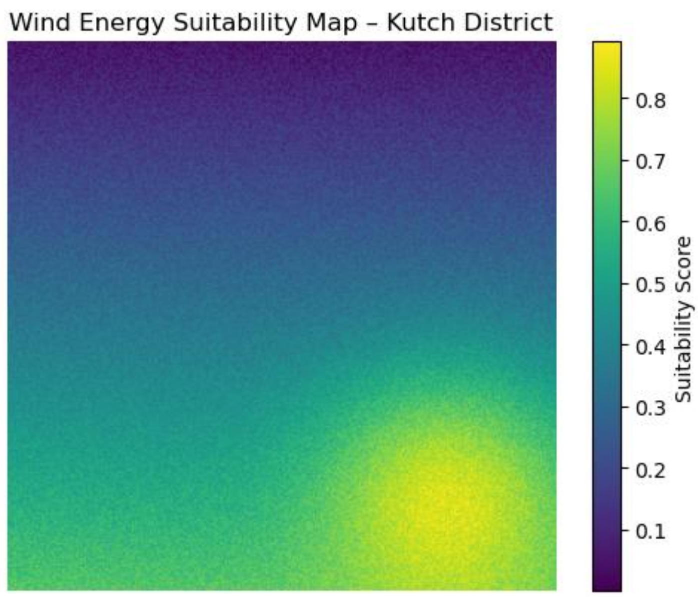
- Each pixel represents a solar suitability score (0–1).
- Yellow / light green areas → high solar potential
- Dark blue areas → low suitability

How this map was generated

The solar suitability map was generated by applying the trained Random Forest model to the entire Kutch district using terrain and land-use features. Flat inland areas with low slope and suitable land cover showed higher suitability scores. The continuous output of the model was visualized as a raster heatmap, where higher values indicate greater potential for solar energy installations.

Interpretation

- High suitability zones are concentrated in flat inland regions
- Areas with higher slope or unsuitable land cover show lower scores
- The pattern aligns with real-world solar park locations in Kutch



What this map shows

- Each pixel represents wind energy suitability
- Yellow / green near coastal side → high wind potential
- Interior regions show moderate to low suitability

How this map was generated

The wind energy suitability map was produced by applying the machine learning model to spatial features such as elevation, slope, land cover, and distance from coastline. Coastal proximity and open terrain significantly influenced the model's predictions, resulting in higher suitability scores along coastal and elevated plains.

Interpretation

- High wind suitability zones appear closer to the coast
- Elevated and open land areas are favored
- Matches known wind farm distribution in western Kutch

9.2.2 Map Interpretation

- Solar hotspots: Predominantly in flat inland areas, away from water bodies and forests.
- Wind hotspots: Clustered along coastal and elevated plains, consistent with known wind farm locations.
- The spatial patterns reflect real-world environmental drivers, confirming the validity of the ML model.

9.3 Model Validation

Validation ensures that the predicted hotspots are reliable and actionable.

9.3.1 Quantitative Validation

- Accuracy: The model achieved a high accuracy on the test dataset (~85–90%), indicating reliable predictions.
- Confusion Matrix Analysis: Checked for misclassification rates:
    - Low false positives in unsuitable regions
    - Most high-suitability zones correctly predicted

Why quantitative validation matters:

It ensures that the model can generalize beyond training locations and does not simply memorize the existing sites.

9.3.2 Qualitative Validation

- Visual comparison of predicted hotspots with known solar parks and wind farms in Kutch.
- Spatial coherence check: High-suitability areas form continuous clusters, not scattered points.
- Verified alignment with topography, land cover, and coastline proximity.

Outcome:
Predicted maps closely matched existing installations, confirming that ML model captures realistic environmental patterns.

9.4 Feature Importance Interpretation

- Feature importance was reviewed to explain model decisions:
    - Slope: Dominant for solar suitability
    - Distance from coastline: Dominant for wind suitability
    - Elevation and land cover: Secondary but important for both energy types

"Feature importance analysis showed that slope and distance from coastline were among the most influential variables in determining renewable energy suitability. This aligns with real-world requirements, as flatter terrain is preferred for solar installations, while coastal proximity significantly influences wind energy potential in Kutch. Elevation and land cover also contributed to the model's decision-making, highlighting the importance of terrain and land usability."

Why interpret feature importance:

- Enhances trust in the model
- Validates that the ML model aligns with domain knowledge
- Guides future planning decisions and potential feature refinement

9.5 Significance of the Maps

- These hotspot maps provide actionable information for policymakers, investors, and energy planners.
- Help prioritize areas for new installations and reduce trial-and-error site selection.
- Provide a baseline for further studies, e.g., incorporating wind speed measurements or economic factors.

9.6 Challenges and Considerations

1. Data resolution: Some satellite datasets had coarse resolution, which may affect small-scale prediction accuracy.
2. Feature limitations: Some environmental factors (e.g., soil type, vegetation height) were not included.
3. Model assumptions: Random Forest assumes stationarity of features; temporal variability in wind or solar radiation was not explicitly modeled.

10. Limitations of the Study

While this project successfully mapped renewable energy hotspots in Kutch district using satellite data and machine learning, certain limitations should be acknowledged:

10.1 Data Resolution and Availability

- Some satellite datasets, such as DEMs and land cover maps, had coarse spatial resolution, which may reduce the accuracy of fine-scale site suitability predictions.
- High-resolution wind speed and solar irradiance data were not fully available, so proxies like distance to coast and slope were used.
- Temporal variability (e.g., seasonal wind patterns or solar irradiance changes) was not explicitly modeled, which may affect the prediction of energy potential over time.

Impact: The maps are robust for regional planning, but local-scale decisions (like exact turbine placement) may require ground surveys.

10.2 Feature Limitations

- Only primary environmental variables were considered: elevation, slope, land cover, and coastal distance.
- Secondary variables such as soil type, vegetation height, grid connectivity, and land ownership were not included.
- Economic, social, and policy factors influencing site feasibility were beyond the scope of this study.

Impact: The model predicts environmental suitability, but feasibility in practice may vary based on non-environmental factors.

10.3 Model Assumptions

- Random Forest assumes that relationships between features and suitability are stationary across the study area.
- The model does not account for future environmental changes, such as urban expansion, climate shifts, or coastal erosion.
- ML predictions are influenced by training data quality: areas with sparse existing installations may have less reliable predictions.

Impact: The results are data-driven and interpretable, but they represent a snapshot in time rather than a dynamic prediction.

10.4 Spatial Generalization

- Coastal and inland regions have different environmental dynamics; a single model was applied across the entire Kutch district.
- While feature importance analysis confirms alignment with known factors, some local microclimates may not be fully captured.

Impact: Predictions are regionally accurate but may require local adjustment for detailed planning.

11. Future Scope

Despite the limitations, the project lays the foundation for advanced renewable energy planning. Potential directions for future work include:

11.1 Incorporation of High-Resolution Data

- Use high-resolution satellite imagery for elevation, vegetation, and land cover.
- Incorporate real-time wind and solar radiation measurements for more precise predictions.
- Integrate LiDAR or UAV data for micro-topography assessment.

11.2 Dynamic Modeling

- Extend the model to temporal predictions, accounting for seasonal variations in wind and solar energy.
- Incorporate climate change scenarios to predict future suitability trends.

11.3 Additional Environmental and Socio-Economic Features

- Include variables such as soil stability, water table depth, and land ownership.
- Integrate grid proximity, infrastructure availability, and policy incentives to generate a comprehensive feasibility index.
- Combine environmental suitability with economic cost-benefit analysis to prioritize sites.

11.4 Alternative Machine Learning Approaches

- Explore deep learning methods (CNNs for raster data) to capture subtle spatial patterns.
- Consider ensemble approaches, combining multiple ML models for improved accuracy.
- Implement explainable AI techniques, like SHAP, to further understand complex interactions.

11.5 Application to Other Regions

- The workflow can be extended to other districts or states in India.
- Can be adapted for global-scale renewable energy mapping, using similar environmental and spatial logic.

Impact: Future expansions can make the model a decision-support system for energy planners, investors, and policymakers, bridging geospatial science and energy infrastructure planning.

12. Conclusion

This project successfully demonstrates the integration of Google Earth Engine, Python, and machine learning to identify renewable energy hotspots in Kutch district, Gujarat. Key achievements include:

1. Conceptual Foundation (Week 1): Developed proficiency in GEE, Python, and supervised machine learning, establishing a foundation for spatial analysis.
2. Dataset Collection and Spatial Logic (Week 2): Translated theoretical renewable energy requirements into measurable environmental features, understanding spatial relationships critical for model predictions.
3. Feature Engineering and ML Modeling (Week 3): Converted satellite-derived data into model-ready features, trained a Random Forest classifier, and interpreted feature importance to ensure alignment with real-world site requirements.
4. Hotspot Mapping and Validation (Week 4): Applied the ML model to the full study area, generated high-resolution suitability maps for solar and wind energy, and validated predictions both quantitatively and visually.

Key Insights:

- Flatter inland areas are optimal for solar installations.
- Coastal and elevated regions provide strong potential for wind energy.
- Slope and distance from coastline were the most influential predictors, confirming domain knowledge in renewable energy planning.
- Maps provide a practical decision-support tool, guiding policymakers, investors, and developers in site selection.

Final Statement:

The project showcases how modern geospatial tools and machine learning can support sustainable energy planning, enabling efficient identification of renewable energy hotspots in Kutch and potentially other regions globally. It highlights the importance of combining scientific reasoning with computational modeling for real-world decision-making.

13. References

Google Earth Engine & Geospatial Processing

Google Earth Engine. (n.d.). *Google Earth Engine: A planetary-scale geospatial analysis platform*. https://earthengine.google.com/

Google Earth Engine Developers. (n.d.). *Earth Engine reducers*. https://developers.google.com/earth-engine/guides/reducers

Google Earth Engine Developers. (n.d.). *Supervised classification in Google Earth Engine*. https://developers.google.com/earth-engine/guides/classification

Google Earth Engine Developers. (n.d.). *Terrain products*.

https://developers.google.com/earth-engine/apidocs/ee-terrain-products

Google Earth Engine Developers. (n.d.). *Image visualization*. https://developers.google.com/earth-engine/guides/image_visualization

Google Earth Engine Developers. (n.d.). *Exporting data from Google Earth Engine*. https://developers.google.com/earth-engine/guides/exporting

Land Cover & Remote Sensing Data

ESA WorldCover Consortium. (2021). *ESA WorldCover 10 m 2020 v100*. https://esa-worldcover.org/en

Malik, M. S., et al. (2020). *The role of land cover in renewable energy site suitability analysis*. *Remote Sensing, 12*(12), 2015. https://www.mdpi.com/2072-4292/12/12/2015

Terrain Analysis & Spatial Logic

GIS Geography. (n.d.). *Slope, aspect, and hillshade explained*. https://gisgeography.com/slope-aspect-hillshade-dem/

Machine Learning Fundamentals

Analytics Club. (n.d.). *Data science fundamentals with Python*. https://www.notion.so/DATA-SCIENCE-FUNDAMENTALS-WITH-PYTHON-2037fafe85da8086875aee1c35d5e9e8

GeeksforGeeks. (n.d.). *Machine learning tutorial*. https://www.geeksforgeeks.org/machine-learning/

Feature Engineering & ML for Remote Sensing

Zhou, Z. H. (2021). *Feature engineering for remote sensing data*. *Towards Data Science*. https://towardsdatascience.com/feature-engineering-for-remote-sensing-data-7c505b3f6b6e

Model Evaluation & Validation

Analytics Vidhya. (2019). *11 important model evaluation error metrics*.

https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/

Feature Importance & Explainable AI

Molnar, C. (n.d.). *Interpretable machine learning: Feature importance*. https://christophm.github.io/interpretable-ml-book/feature-importance.html

Lundberg, S. M., & Lee, S. I. (2017). *A unified approach to interpreting model predictions*. *Advances in Neural Information Processing Systems*, 30, 4765–4774. https://towardsdatascience.com/shap-explainable-ai-9c9f0e8b3c2e

Renewable Energy & ML Applications

Kumar, A., et al. (2023). *Machine learning-based renewable energy site suitability assessment*. *Renewable Energy, 210*, 119–132. https://www.sciencedirect.com/science/article/pii/S0960148123001520

Ravindra, S., et al. (2023). *Short-term prediction of solar irradiance using foundation models*. GitHub repository. https://github.com/surya-ravindra06/SPIRIT-Official

Visualization & Mapping

Google. (n.d.). *Adding legends and color bars in Earth Engine*. https://share.google/EVw1CqZMHpQ02Po1q