Solar Power Plant Location Classifier: Executive Summary

Utkarsh Agrawal, Mansi Bezbaruah, Harsha Hampapura, Akshay Khadse, Meysam Motaharfar

GitHub: https://github.com/harshahampapura/SolarFarmPrediction.git

Overview and Motivation

Facing climate challenges and the need for sustainable growth, the global community is at a crucial point. To address this, we must accelerate the adoption of renewable energy, with solar power being the most promising source due to its high power generation efficiency (~200 W/m^2). The limitations of rooftop solar highlight the **urgency of prioritizing utility-scale solar farms** (typically 1 MW or larger) to meet rising energy demands. Governments, businesses, and communities globally are recognizing the significance of solar farms, with California witnessing a 100% growth in utility-scale solar farms in the last 5 years alone. Therefore, optimizing locations for these solar farms is crucial for efficiency and output maximization. Traditional site selection methods involve manual, time-intensive processes or expert land surveys. Our innovative project uses geospatial data analytics and machine learning algorithms to **streamline the site selection process**. Our predictive models consider factors like solar irradiance, topography, land use, and proximity to cities. The result is a recommendation system for potential solar farm construction sites, significantly reducing time and resources spent on research. This tool is **indispensable for stakeholders in the pursuit of sustainable energy solutions.**

Stakeholders: Energy and utility firms, landholders leasing their land to solar energy farms, state and local governments, green energy companies.

Key Performance Indicators (KPIs): (Efficiency Index) How efficiently does the model predict suitable locations compared to actual installations?

Approach

In seeking ideal utility-scale solar farm locations, we gathered data on construction cost/viability, energy output, and local electricity demand. We focused on five key features: modeled annual average AC electricity output, land cover, elevation, slope, and proximity to the nearest city by road. The modeled AC output indicates the maximum solar energy harnessable at a given location for a 1MW capacity farm, factoring in typical losses and weather conditions, sourced from the National Renewable Energy Laboratory's solar model. This approach negates the need for separate training on variables like solar irradiance, temperature, and cloud cover. Using the annual average accounts for variations in solar output over shorter periods. Topographical features and city proximity inform us about the landscape and potential electricity demand. Our dataset includes these **five distinct features across 971 instances**. After data cleaning, we applied an ensemble of machine learning models, including deep learning, for robustness. This mix of algorithms, from simple to complex, and a voting method across weak learner classifiers, aimed to enhance overall accuracy. Our classification goal led to the selection of **Logistic Regression, XGBoost Classifiers, and Deep Learning with classification layers**, alongside **Support Vector Machines and Decision Trees,** balancing model interpretability and linearity.

Challenges Faced

Challenges in our project included a lack of reliable data on local electricity demand, land and labor prices, and maintenance costs. Additionally, we encountered solar farms located on water bodies, requiring manual exclusion of such locations from our land-based solar farm analysis. We also aimed to build a website for easy classification of potential solar farm sites. However, we faced a number of roadblocks implementing this and intend to complete this in the future.

Results

As stated above, our goal was to create a model that can **classify if a given location is suitable for a utility-scale solar farm**. After carefully collecting and cleaning the data we were able to train models that **classify existing solar farms with 95% accuracy.** This, combined with the fact that our models had similar accuracy for training and test datasets suggests that our models are robust and do not suffer either high bias or variance. Moreover, we also found that **by training on older solar farm data, we could predict the locations of newly-built solar farms > 95% of the time**. This serves as an important independent test of our models and demonstrates that our models generalize well to newer datasets. A key outcome is quantifying geographic and climatic factors' impact on solar farm viability. Surprisingly, we find that all the five features we have explored contribute equally to a location being ideal for a solar farm.

Future Work

There are various ways in which we can further improve our model.

1. ***Adding Features***: There are other features that we would like to include in the model. But, we were unable to do so for various reasons. We could include commercial data such as land cost, human labor and transportation cost, profit generated, etc into the model. Moreover, we would like to quantify measures like government support, for example, by introducing a metric and checklist for local laws and incentives towards solar energy. We also wish to build a website to easily test potential solar sites.
2. ***Including more locations***: Due to time constraints, we had to restrict our data collection to the state of California. We could further train and test our model on locations outside California. This would increase the accuracy of our models and would provide us with more points to test the efficiency index on.