Solar Farm Location Scout and ROI Calculator based on U.S. Weather Data

Edward Seymour
25 September, 2022

As a data scientist, I will use U.S weather and irradiation data to create a model able to estimate solar energy generation at points across the country, such that commercial users will be able to locate optimal locations for solar installations and residential customers will be able to calculate the amount of money they could save on energy per year for a given amount of solar panels.

The U.S. is rapidly shifting to methods of renewable energy generation in response to the growing impacts of man-made climate change. The most popular, by both residential and commercial customers, is solar. Residential customers are able to sell excess energy back to their energy provider and commercial customers are able to provide a service that isn't beholden to the fluctuating prices of natural gas, coal, or oil. This makes it a smart choice for both parties. The country is currently investing heavily in solar energy and potential customers will need the information necessary to make the right financial decision. Using weather and irradiation data for the U.S., I will develop a model to identify the optimal locations for solar installations, as well as a ROI calculator for residential customers using location-based electricity prices and average home electricity use.

For this project, I will focus only on U.S. solar energy prospects, no other renewables. The model must be able to locate optimal locations for solar installations for business customers, as well as calculate the estimated amount of power generated over the year for that location for a given area of solar panels. For residential customers, the model needs only to calculate the ROI, using the same method to estimate the amount of power generated over the year.

The model will be based on weather data and solar radiation data. As climate change worsens, an ROI calculated for a specific weather and irradiation pattern might no longer be applicable to the same customer that it was previously. It is important for users of the model to return to the model on a repeated basis to monitor their investment's outlook. In the same vein, the model needs to work on continuously updated data.

The stakeholders on this project are established energy companies and start-ups looking to invest in solar installations and residential customers looking to decrease their energy bill and carbon footprint.

The key data sources for this project are:

- Photovoltaic (PV) Solar Panel Energy Generation Data
    - Data from UK Power Networks, available on London Datastore. The dataset includes weather data and solar energy generation at 30 sites over more than a year. This data will be used to train the model to predict energy generated based on available weather data.
    - https://data.london.gov.uk/dataset/photovoltaic--pv--solar-panel-energy-generation-data
- NSRDB - National Solar Radiation Database
    - Large dataset from the Open Energy Initiative. This dataset will be used to find optimal locations in the U.S. for solar installations based on the factors found by the model. This data will also be used to calculate ROI for specific locations for residential customers.
    - https://data.openei.org/s3_viewer?bucket=nrel-pds-nsrdb&prefix=v3%2F