Chad, Morgan and Dominique

Solar PV Yield and Loss

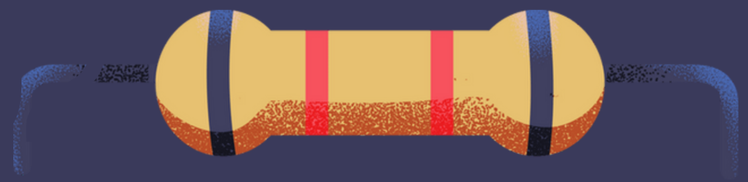
Extract and predict renewable energy performance and losses to improve resource allocation

Data Science #1448

Project Relevance & Significance

Solar PV is an economical approach to address loadshedding in South Africa.

It combines reduction in overall grid demand with a renewable and sustainable approach to power generation.



Data Science Objective

Analyse solar PV system performance countrywide to spot potential performance trends.

Identify and label events of power loss.

Develop machine learning models to predict and compare future solar PV yields based on historical data.

The Approach

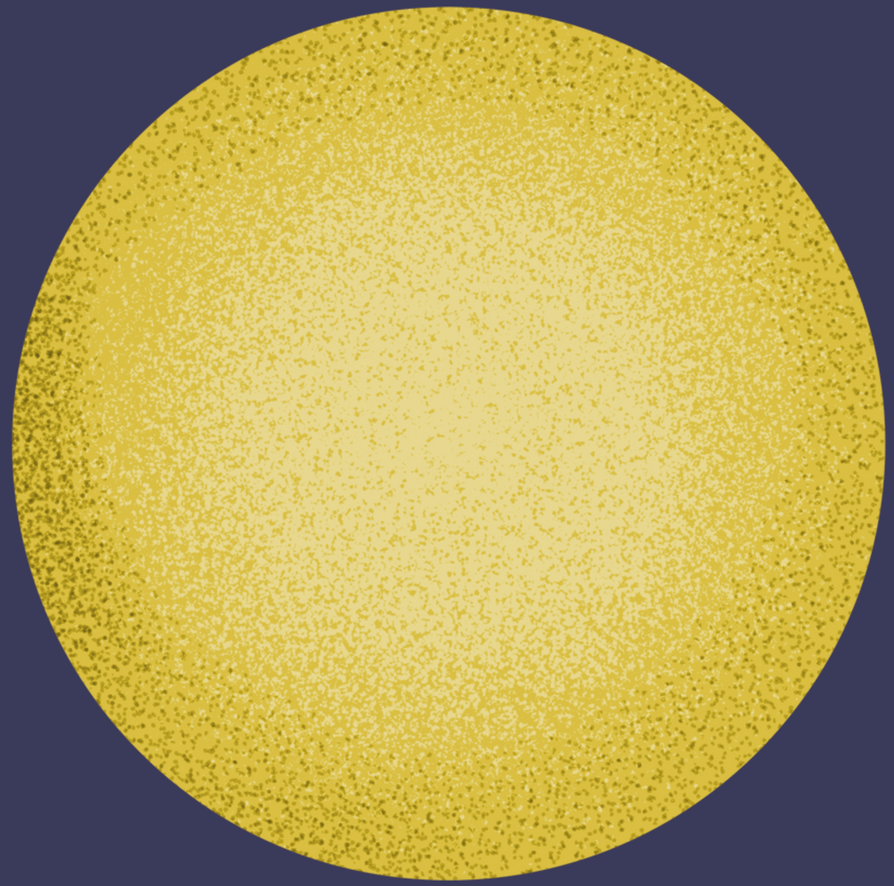
Combine solar PV meter and satellite weather data to create a useable data set.

Perform feature engineering and preprocessing.

Analyse and visualise data to spot trends.

Create power loss event detection feature.

Find and implement the most suitable prediction model.

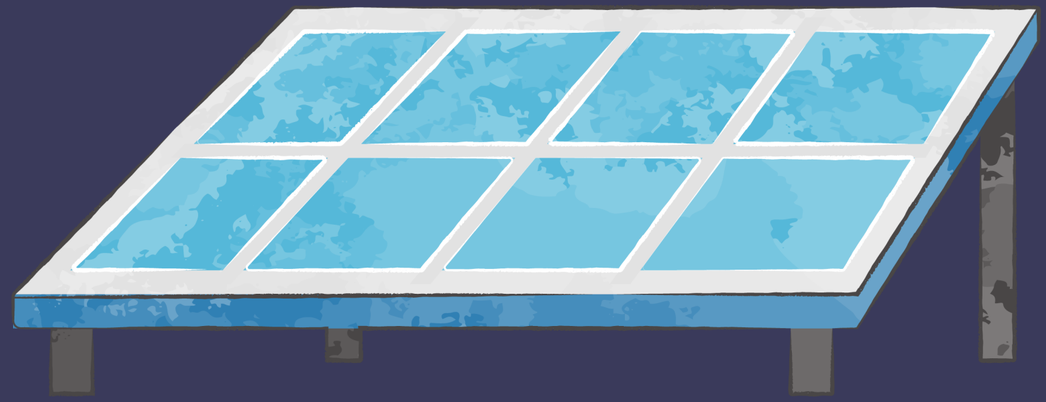


Challenges & Work Done

Extensive data cleaning and correction from different data sources and types.

Multiple machine and deep learning approaches attempted - MSE often in the 1000kWh range.

Application created displaying notable features and the target variable - total power produced.



The Final Model

Tree-based Pipeline Optimization Tool (TPOT) to the rescue:

- RobustScaler
- Cross-validated Lasso using LARS (least angle regression) algorithm

MSE error of 2.669e-26!

RESULTS LIVE DEMO - STREAMLIT

Moving Forward

Creating a data set that has systems with more varied characteristics as well as data going back further in time.

Train the model on this expanded data to hopefully be more robust.

Build a power loss predictor model.

Include financial reporting information such as ROI and tariff savings.

Provide Cost Benefit Analysis metrics.