Solar Panel Regression

AGENDA

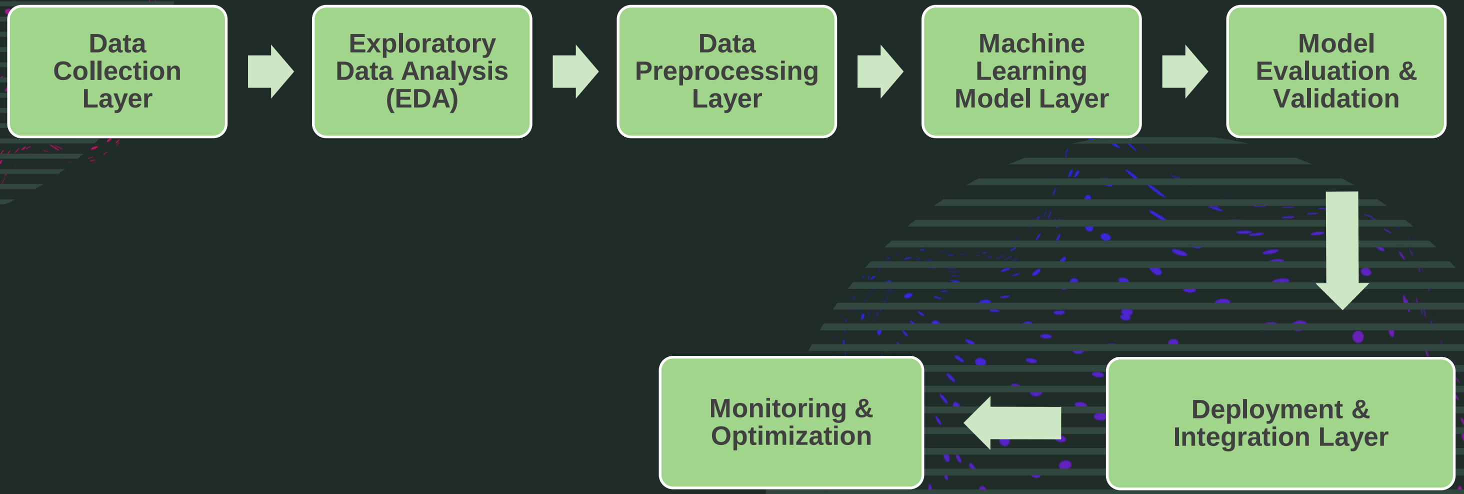
Solar Panel Performance Prediction Using Machine Learning

Prepared By : Group 3

GOKILA G
REDDYCHERLA ARUNKUMAR
JUDE S F
SHAHID HUSSAIN S
NISCHITHA R
MOHAMMED THOUSIF

Contents

Project Architecture



Introduction

- Solar panel regression helps predict energy output by analyzing factors like sunlight, temperature, and dust. Accurate forecasting optimizes performance, reduces costs, and improves maintenance efficiency.
- Machine learning models use historical and real-time data to enhance prediction accuracy. This enables better decision-making, maximizing energy production and ensuring sustainable solar power management.
- Advanced regression techniques, such as deep learning and ensemble methods, further improve prediction reliability, helping large-scale solar farms and smart grids operate more efficiently.

DATA SET DETAILS

```
<class 'pandas.core.frame.DataFrame'>
Index: 2811 entries, 0 to 2919
Data columns (total 10 columns):
 #   Column                      Non-Null Count  Dtype  
---  ------                      --------------  -----  
 0   distance_to_solar_noon      2811 non-null   float64
 1   temperature                 2811 non-null   int64  
 2   wind_direction              2811 non-null   int64  
 3   wind_speed                  2811 non-null   float64
 4   sky_cover                   2811 non-null   int64  
 5   visibility                  2811 non-null   float64
 6   humidity                    2811 non-null   int64  
 7   average_wind_speed_period   2811 non-null   float64
 8   average_pressure_period     2811 non-null   float64
 9   power_generated             2811 non-null   int64  
dtypes: float64(5), int64(5)
memory usage: 241.6 KB
```

Data Preprocessing And EDA

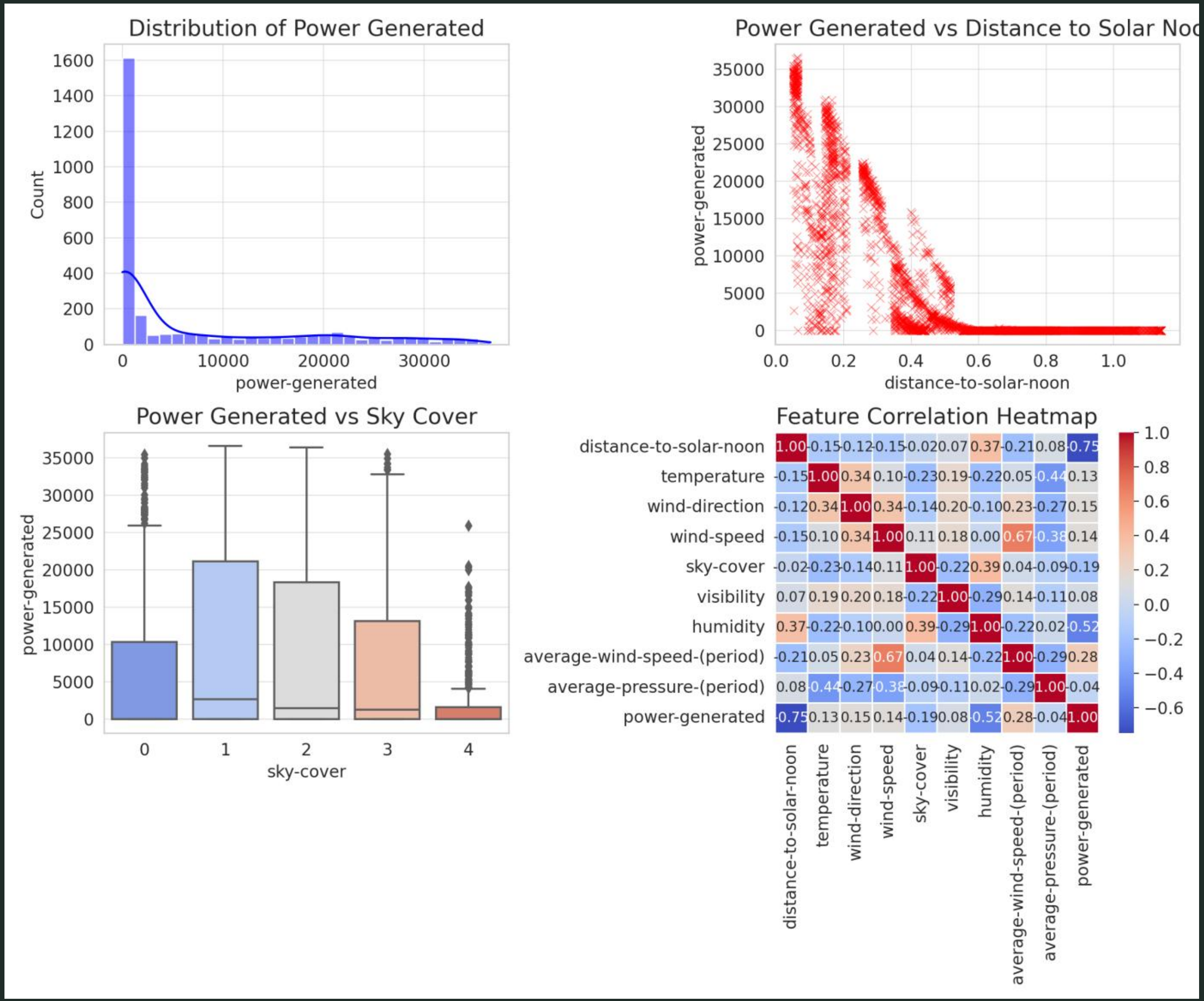
- **Missing Values:** Only average-wind-speed-(period) has one missing value (can be filled with the mean).
- **Correlation with Power Generated:**

  Strong negative correlation with distance-to-solar-noon (-0.75) → Power generation is highest around solar noon.

  Moderate negative correlation with humidity (-0.52) → More humidity reduces power generation.

  Weak positive correlation with wind-speed (0.14) and temperature (0.13).

  Low correlation with average-pressure-(period) (-0.04).

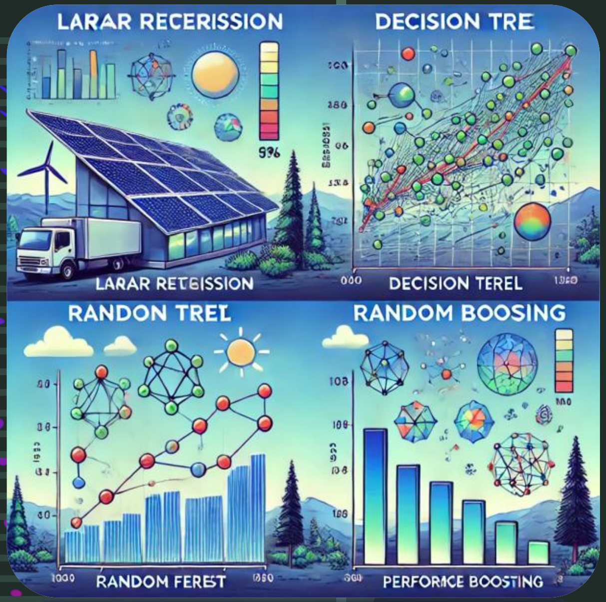


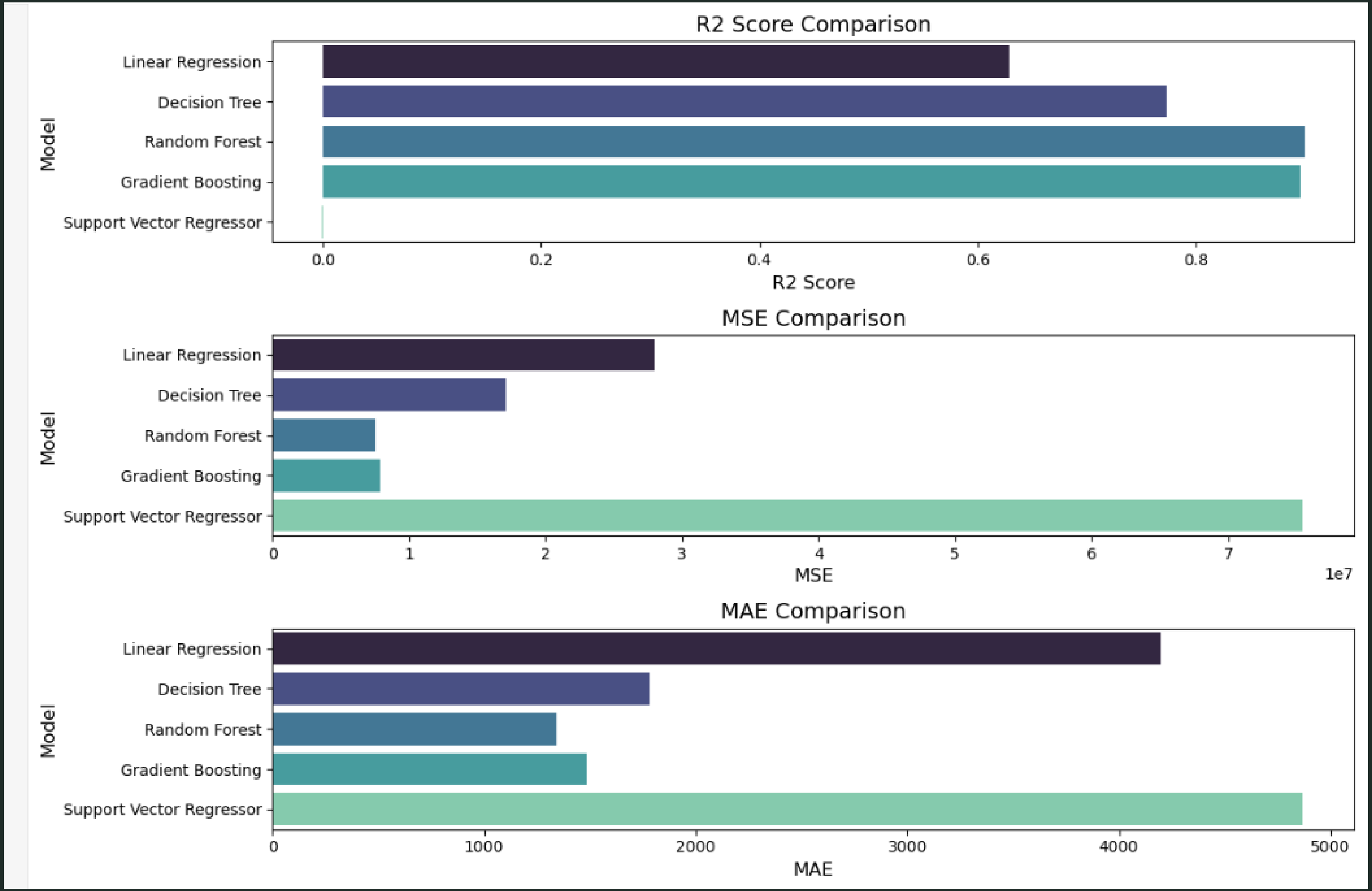
- **Power Generation Distribution** – Most values are low, but some high peaks exist.
- **Power vs. Distance to Solar Noon** – Strong negative relationship, confirming peak generation at noon.
- **Power vs. Sky Cover** – Higher cloud cover generally reduces power generation.
- **Feature Correlation Heatmap** – Highlights key relationships, such as negative correlation with humidity and distance to solar noon.

Model Developing

Splitting data into **training**, **validation**, and **test sets** ensures effective model training, tuning, and evaluation, with each set serving a specific purpose in the process.

- Linear Regression**: R^2 = 0.63, MAE = 4200
- Decision Tree**: R^2 = 0.77, MAE = 1783
- Bagging Regression**: R^2 = 0.88, MAE = 1424
- Gradient Boosting**: R^2 = 0.90, MAE = 1487
- Random Forest (Best Model)**: R^2 = 0.90, MAE = 1343 ✅
- Support Vector Regressor**: Poor performance (R^2 = -0.001)




R2 Score Comparison
Model
Linear Regression
Decision Tree
Random Forest
Gradient Boosting
Support Vector Regressor
0.0
0.2
0.4
0.6
0.8
R2 Score
MSE Comparison
Model
Linear Regression
Decision Tree
Random Forest
Gradient Boosting
Support Vector Regressor
0
1
2
3
4
5
6
7
1e7
MSE
MAE Comparison
Model
Linear Regression
Decision Tree
Random Forest
Gradient Boosting
Support Vector Regressor
0
1000
2000
3000
4000
5000
MAE

Model Evaluation

- **R^2 Score** – Measures how well the model explains variability in solar power generation. A higher score indicates better performance.
- **Mean Squared Error (MSE)** – Represents the average squared difference between actual and predicted values. Lower MSE means better accuracy.
- **Mean Absolute Error (MAE)** – Measures the average absolute error between predictions and actual values, providing insight into model accuracy.
- **Model Comparison** – Evaluating models like Linear Regression, Decision Tree, Random Forest, and Gradient Boosting helps select the best one.

Model Deployment

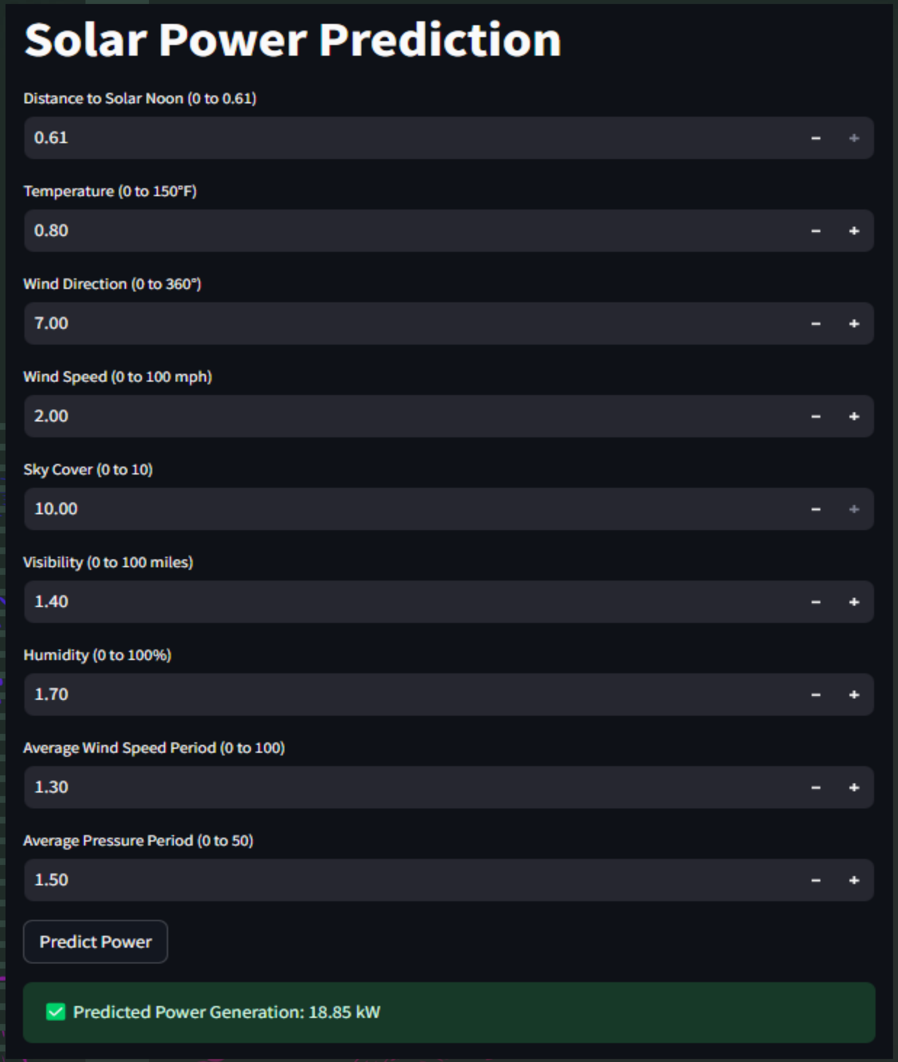
- Model deployment for Solar Panel Regression enables real-time predictions of solar power generation based on environmental factors. Streamlit provides an interactive interface for users to input data and visualize predictions easily.

```python
import pickle
import pandas as pd
import streamlit as st
```

```python
# Load the trained model
model = pickle.load(open("solar_power_model.pkl", "rb"))

# Streamlit UI
st.title("Solar Power Prediction")

# Define valid input ranges
valid_ranges = {
    "distance_to_solar_noon": (0, 0.61),
    "temperature": (0, 150),
    "wind_direction": (0, 360),
    "wind_speed": (0, 100),
    "sky_cover": (0, 10),
    "visibility": (0, 100),
    "humidity": (0, 100),
    "average_wind_speed_period": (0, 100),
    "average_pressure_period": (0, 50),
}
```

```python
# Collect user input
distance = st.number_input("Distance to Solar Noon (0 to 0.61)", min_value=0.0, max_value=0.61, step=0.01)
temperature = st.number_input("Temperature (0 to 150°F)", min_value=0.0, max_value=150.0, step=0.1)
wind_direction = st.number_input("Wind Direction (0 to 360°)", min_value=0.0, max_value=360.0, step=1.0)
wind_speed = st.number_input("Wind Speed (0 to 100 mph)", min_value=0.0, max_value=100.0, step=0.1)
sky_cover = st.number_input("Sky Cover (0 to 10)", min_value=0.0, max_value=10.0, step=1.0)
visibility = st.number_input("Visibility (0 to 100 miles)", min_value=0.0, max_value=100.0, step=0.1)
humidity = st.number_input("Humidity (0 to 100%)", min_value=0.0, max_value=100.0, step=0.1)
average_wind_speed = st.number_input("Average Wind Speed Period (0 to 100)", min_value=0.0, max_value=100.0, step=0.1)
average_pressure = st.number_input("Average Pressure Period (0 to 50)", min_value=0.0, max_value=50.0, step=0.1)

# Collect user inputs into a dictionary
user_inputs = {
    "distance_to_solar_noon": distance,
    "temperature": temperature,
    "wind_direction": wind_direction,
    "wind_speed": wind_speed,
    "sky_cover": sky_cover,
    "visibility": visibility,
    "humidity": humidity,
    "average_wind_speed_period": average_wind_speed,
    "average_pressure_period": average_pressure,
}

# Predict button
if st.button("Predict Power"):
    # Validate user input
    invalid_inputs = [
        (name, value) for name, value in user_inputs.items()
        if not (valid_ranges[name][0] <= value <= valid_ranges[name][1])
    ]

    if invalid_inputs:
        for name, value in invalid_inputs:
            st.error(f"❌ {name.replace('_', ' ').title()} must be between {valid_ranges[name][0]} and {valid_ranges[name][1]}. You entered: {v
    else:
        # Prepare data for prediction
        data = pd.DataFrame([list(user_inputs.values())], columns=user_inputs.keys())

        # Make prediction
        prediction = model.predict(data)[0]
        st.success(f"✅ Predicted Power Generation: {prediction:.2f} kW")
```

Streamlit Interface



Challenges in Solar Panel Regression

- Variability in Weather Conditions – Factors like cloud cover, temperature, and humidity affect solar power generation, making predictions challenging.
- Non-Linear Relationships – The impact of sunlight intensity and temperature on power output is complex, requiring advanced regression models.
- Data Quality and Availability – Missing, inconsistent, or limited historical data can reduce model accuracy and reliability.
- Feature Selection – Identifying the most relevant environmental and system parameters is crucial for improving prediction accuracy.
- Model Generalization – A model trained on one location's data may not work well in other regions with different weather conditions.

Thank you