**Case Study**

**Predicting Renewable Energy Production Using Historical Data and Weather Patterns**

**Introduction**

Renewable energy sources, such as solar and wind, are increasingly critical for sustainable energy systems. However, their production can be highly variable and dependent on weather conditions. This case study focuses on using data science techniques to predict renewable energy production, utilizing historical data and weather patterns to enhance the efficiency and reliability of these sources.

**Objectives**

* **Collect and preprocess data:** Gather historical energy production and weather data.
* **Feature engineering:** Create relevant features for the model.
* **Model selection and training:** Use machine learning algorithms to predict energy production.
* **Model evaluation:** Assess model performance using appropriate metrics.
* **Visualization:** Generate visualizations to understand the model and the data.

**Data Collection**

For this study, synthetic data representing two locations was created. The data includes daily records from January 1, 2020, to January 1, 2023, covering weather parameters (temperature, wind speed, solar radiation, and humidity) and energy production values.

**Data Preprocessing**

* **Merge Datasets:** Combine historical energy production data with weather data based on date and location.
* **Handle Missing Values:** Ensure no missing values to maintain data integrity.
* **Feature Engineering:** Extract additional features like month and year from the date to capture seasonal variations.

**Feature Engineering**

One-hot encoding was used for the location feature to convert categorical data into numerical format suitable for machine learning models. Additionally, the date was split into month and year to account for seasonal and yearly trends.

**Model Selection and Training**

A Random Forest Regressor was chosen due to its robustness and ability to handle complex interactions between features. The data was split into training (80%) and testing (20%) sets. The model was trained on the training set and evaluated on the testing set.

**Model Evaluation**

The model's performance was assessed using Mean Squared Error (MSE) and R^2 score. These metrics provide insights into the accuracy and explanatory power of the model. The model achieved an MSE of approximately 918.76 and an R^2 score of around 0.78, indicating a good fit for the data.

**Visualization**

Several visualizations were generated to better understand the model and the data:

* **Actual vs Predicted Values:** A line plot comparing actual energy production with predicted values showed how closely the model's predictions matched the real data.
* **Feature Importances**: A bar plot displaying the importance of each feature in the model, highlighting which weather parameters most influenced energy production.
* **Pair plot:** Visualizing relationships between features and energy production helped in understanding the correlations between variables.
* **Distribution Plot:** Displaying the distribution of energy production values provided insights into the central tendency and spread of the data.
* **Time Series Plot:** Energy production over time for different locations illustrated trends and seasonal patterns.

**Conclusion**

The case study demonstrates that machine learning models, particularly Random Forest, can effectively predict renewable energy production based on historical data and weather patterns. The approach not only enhances the reliability of renewable energy sources but also aids in better planning and utilization. Future work can focus on incorporating real-world data, refining model parameters, and exploring other machine learning techniques to further improve prediction accuracy.