Analyzing the Impact of Weather Patterns on Renewable Energy Production

Background and Motivation:

The worldwide transition to renewable energy sources is crucial for mitigating carbon emissions and addressing climate change. China, as the world's leading producer of renewable energy, is at the forefront of this transition. Among various renewable sources, solar and wind power are pivotal due to their abundant availability and low environmental impact. However, the unpredictable and intermittent nature of solar and wind energy presents significant challenges for their integration into the power grid and optimization of power systems. Precise forecasting of solar and wind power generation is vital for maintaining grid stability, balancing power supply and demand, and facilitating informed decision-making in energy planning and policy development. This forecasting is particularly crucial in China, where the rapid expansion of renewable energy capacity necessitates advanced predictive models to manage the complexities of the energy grid effectively.

Problem Definition

- Objective:

The project aims to develop a predictive model to forecast solar and wind power generation using on-site data collected from six wind farms and eight solar stations in China. The model will cover different climate zones and geographic locations, providing a comprehensive understanding of the relationship between weather conditions and renewable energy output.

- Importance:

Accurate forecasting of renewable energy generation is crucial for day-ahead power scheduling, optimizing the power system, and allowing for higher renewable energy penetration into the grid. It addresses the intermittency and fluctuation characteristics of solar and wind energy, which are major challenges in renewable energy generation.

- Input variables (weather conditions):
    - Wind speed is the speed of the wind, which directly influences the output of wind turbines.
    - Wind Direction: The direction from which the wind is blowing, affecting the performance of wind turbines.
    - Total Solar Irradiance (TSI): The total amount of solar radiation received by the solar panels is critical for solar power generation.
    - Air Temperature: The ambient temperature, which can affect the efficiency of both solar panels and wind turbines,.

    - Atmospheric pressure is the pressure of the atmosphere, which can influence wind patterns and solar irradiance.
    - Relative Humidity: The amount of moisture in the air, which can impact solar panel efficiency.
- Target Variable:
    - Energy Output (Power in MW): The amount of energy produced by the solar panels and wind turbines, measured in megawatts (MW). This is the target variable that the model aims to predict based on the input weather conditions.

By specifying these input and target variables, the project aims to develop a model that can accurately predict the energy output of solar and wind power generation systems based on various weather conditions.

Data Collection and Refinement

- **Data Sources:**
    - The dataset comprises over two years (2019–2020) of power generation and weather-related data collected at 15-minute intervals from six wind farms and eight solar stations in China. The data includes key parameters such as wind speed, wind direction, total solar irradiance, air temperature, atmospheric pressure, and relative humidity. These stations are strategically located across various climate zones and terrains, including deserts, mountains, and plains, to ensure diverse and representative data.
    - In addition to the on-site data collected from six wind farms and eight solar stations in China, this project will also utilize the CarbonMonitor-Power dataset to provide a broader context of renewable energy generation in China. The CarbonMonitor-Power dataset offers near-real-time global power generation data, including hourly-to-daily time scales for renewable sources such as solar and wind power. With over two million records, this dataset provides valuable insights into the energy mix of China and other countries, highlighting the contributions of renewable energy to the overall power generation.

Wind Farm Dataset

Heading Name	Description
Wind Speed at Height of X Meters (m/s)	The wind speed measured X meters above the ground.
Wind Direction at Height of X Meters (°)	The wind direction measured X meters above the ground, in degrees.

Air Temperature (°C)	The dry-bulb air temperature measured at 1.5 meters above the ground.
Atmospheric Pressure (hPa)	The atmospheric pressure measured at 1.5 meters above the ground.
Relative Humidity (%)	The relative humidity of the air measured at 1.5 meters above the ground.
Power (MW)	The total power generation from the wind farm, measured in megawatts (MW).

Solar Station Dataset

Heading Name	Description
Total Solar Irradiance (W/m²)	The total solar power over all wavelengths per square meter.
Direct Normal Irradiance (W/m²)	The amount of solar radiation received per square meter by a surface perpendicular to the rays.
Global Horizontal Irradiance (W/m²)	The total amount of shortwave radiation received by a surface horizontal to the ground.
Air Temperature (°C)	The dry-bulb air temperature measured at 1.5 meters above the ground.
Atmospheric Pressure (hPa)	The atmospheric pressure measured at 1.5 meters above the ground.
Relative Humidity (%)	The relative humidity of the air measured at 1.5 meters above the ground.
Power (MW)	The total power generation from the solar station, measured in megawatts (MW).

CarbonMonitor-Power dataset (For Background Information)

Heading Name	Description
date	The date when the power generation data was recorded.
country	The country where the power generation data was recorded.

type	The type of power generation source (e.g., Coal, Gas, Oil, Nuclear, Hydroelectricity, Solar, Wind, Other sources).
value	The amount of power generated by the specified source, measured in gigawatt-hours (GWh).
label	Indicates whether the data is a forecast (F) or a normal record (N).

- **Data Refinement:**
    - **Combining Datasets:** Since we have 6 datasets for wind and 8 datasets for solar from different stations, we will combine them to focus on weather conditions before proceeding with further refinement.
    - Missing Data: The missing data include variables that were zero, null, ‘NA’, ‘0.001’, ‘−99’, and ‘–’. We will employ various techniques to handle missing data based on the extent and nature of the missing values. For short-term missing data (less than 10 consecutive time steps), we will use upward/downward completion or linear interpolation to fill in the gaps. For intermittent missing data (10 to 100 consecutive time steps), a moving average method will be applied. For long-term missing data (more than 100 consecutive time steps), the affected samples will be removed from the dataset to maintain data integrity. The dataset should not be adopted if the missing data rate is larger than a specific rate (i.e., 20%) of the total dataset.
    - Outliers:The outliers include weather variables that remained unchanged over a long time, atmosphere values that were equal to zero, and values that were unreasonably high or low. Outliers will be detected using statistical methods, such as identifying data points that lie outside 1.5 times the interquartile range (IQR). These outliers will be examined to determine if they are due to measurement errors or represent genuine fluctuations in renewable energy production. Based on this assessment, outliers will either be corrected or excluded from the dataset. Some outliers may be considered meaningful data points for developing a data-driven forecasting model due to the fluctuated characteristics of renewable energy.

- **Feature Selection:**
    - Correlation Analysis: We will conduct a correlation analysis to identify the weather variables that have the strongest relationship with renewable energy output. Variables such as wind speed, wind direction, total solar irradiance, air temperature, atmospheric pressure, and relative humidity will be considered. The analysis will help us select the most relevant features for inclusion in the predictive model.

- Seasonal and Temporal Patterns: In addition to correlation analysis, we will also examine seasonal and temporal patterns in the data to identify any cyclical trends that could impact renewable energy production. This will enable us to refine our feature selection further and improve the accuracy of our predictive model.

Implementation

- Modeling Approach:
    - Diverse Model Comparison: By developing models like LSTM and Random Forest in addition to GANs, I aim to cover a broad spectrum of predictive capabilities. LSTM networks are well-suited for capturing temporal dependencies in time series data, which is critical for forecasting based on historical weather patterns and their impact on power generation. Random Forest, on the other hand, can effectively handle non-linear relationships and interactions between multiple input features without the need for extensive parameter tuning. This diversity in modeling approaches ensures a comprehensive exploration of the data's predictive capacity.
    - Complexity and Nonlinearity: GANs are particularly adept at capturing complex nonlinear relationships and generating high-dimensional data distributions. This makes them suitable for the task of forecasting power generation based on a multitude of interrelated weather conditions. The adversarial process of training a generator and a discriminator concurrently allows GANs to produce highly realistic forecasts, potentially capturing nuances that simpler models might miss.
    - Parameter Tuning and Optimization: By incorporating techniques like grid search or random search for hyperparameter optimization, I aim to fine-tune each model's parameters to achieve optimal performance. This step is essential to ensure that the models are not underfitting or overfitting the data. Cross-validation will be used to validate the models' performance on different subsets of the data, providing a more reliable estimate of their predictive capabilities.
    - Benchmarking and Accuracy: By comparing the performance of GANs with that of LSTM, Random Forest, and possibly other models, I aim to establish a benchmark for accuracy in renewable energy forecasting. This comparative analysis will help identify the most effective model or combination of models for accurately predicting energy production, thereby informing future forecasting efforts and model development.
- Carbon cost calculation:
    - To assess the environmental impact of our predictive models, we will calculate their carbon cost using the CodeCarbon library. CodeCarbon is an open-source tool that estimates the carbon emissions associated with the computational resources consumed during model training. The calculation will be based on

factors such as energy consumption, the location of the computing resources, and the duration of the model training process.

        - Implementation Steps:
            - Integration of CodeCarbon: We will integrate the CodeCarbon library into our model training pipeline to automatically track and estimate the carbon emissions during the training process.
            - Data Collection: CodeCarbon will collect data on energy consumption and other relevant parameters during model training.
            - Emission Estimation: Based on the collected data, CodeCarbon will estimate the carbon emissions using its built-in emission factors for different regions and energy sources.
            - Reporting: The estimated carbon emissions will be reported and documented as part of the model's evaluation metrics. This will provide insights into the environmental impact of the predictive models and help in making informed decisions regarding their deployment and usage.
        - By calculating the carbon cost of our models, we aim to ensure that our project not only contributes to the advancement of renewable energy forecasting but also aligns with sustainability goals by minimizing the environmental impact of the computational processes involved.

- Model Training and Validation:
    - The dataset will be divided into training and validation sets, with a typical split of 80% for training and 20% for validation. The best performance model will be trained on the training set, where the generator will learn to produce realistic power generation data based on the input weather conditions, and the discriminator will learn to distinguish between real and generated data.
    - The model's performance will be evaluated on the validation set using specific metrics to assess its accuracy and reliability.
- Model Testing and Evaluation:
    - After training and validation, the model will be tested on unseen data to evaluate its generalization capability. This step is crucial to ensuring that the model can accurately predict renewable energy production in different scenarios and conditions.
    - The model's performance will be analyzed across various climate zones and geographic locations to ensure its robustness and applicability in diverse settings.

Expected Outcomes:

- A robust and reliable predictive model capable of accurately forecasting solar and wind power generation based on weather conditions.
- Detailed insights into how different meteorological factors influence renewable energy output, aid in the development of optimization strategies for renewable energy systems and grid integration.
- A valuable tool for energy planners, grid operators, and policymakers to enhance the reliability and efficiency of renewable energy sources in the power grid.

Significance

- Grid Stability: By providing accurate forecasts of renewable energy generation, the model will help in better grid management, reducing power imbalance problems, and ensuring a stable energy supply.
- Demand Response Optimization: The insights gained from the model can be utilized to optimize electricity demand response programs, aligning energy demand with the availability of renewable energy, thus improving the overall efficiency of the power system.
- Sustainability: Accurate forecasting of renewable energy generation contributes to increasing the penetration of renewable sources in the energy mix, reducing reliance on fossil fuels, and promoting environmental sustainability.

Reference

Liu, H., Zhang, Z., Wang, Z., Chen, M., Ding, J., & Cai, Z. (2021). Solar and wind power data from the Chinese State Grid Renewable Energy Generation Forecasting Competition [Data set]. Figshare. https://doi.org/10.6084/m9.figshare.17304221.v4

Zhu, B. et al. CM-Power near-real-time monitoring of global power generation on hourly to daily scales. Figshare https://doi.org/10.6084/m9.figshare.21938102.v2 (2022).