This hackathon combines two separate prediction tasks:

1) **Wind Power Generation Forecasting:** Predicting the total power generated by wind.
2) **Smart Grid Stability Assessment:** Determining the stability of the grid based on power generation, consumption, and other factors.

1) Wind Energy Power Generation:

Input features: Air Temperature, Pressure, Wind Speed

Target variable: Power generated (MW)

This is a regression problem, where you want to predict the power generated based on the input features.

Suitable models: Linear Regression, Decision Tree Regression, Random Forest Regression, Gradient Boosting Regression, Neural Networks (for non-linear relationships).

2) Smart Grid Dataset:

Input features: Power Generated (power_gen_1, power_gen_2, power_gen_3), Price per unit (p1, p2, p3), Power Consumption (C1, C2, C3)
Target variable: 'stability' (categorical - 'stable' or 'unstable')

This is a classification problem, where you want to predict the stability of the system based on the input features.

Suitable models: Logistic Regression, Decision Tree Classifier, Random Forest Classifier, Gradient Boosting Classifier, Support Vector Machines (SVMs), Neural Networks.

Here's a breakdown of the best models for each task:

1. Wind Power Generation Forecasting:

<u>Recommended Model:</u> Random Forest Regression

<u>Reasoning:</u>

- Can handle non-linear relationships between features like Air Temperature, Pressure, and Wind Speed.
- Works well with provided forecasted values for independent variables.
- Generally robust to outliers in data.

2. Smart Grid Stability Assessment:

<u>Recommended Model:</u> Support Vector Machine (SVM) Classification

<u>Reasoning:</u>

- Effective for classifying data points into distinct categories (stable/unstable) based on various power and consumption parameters.
- Handles high-dimensional data (many parameters) from the "Grid folder" data.

<u>Overall Approach:</u>

- Use Random Forest Regression to predict total power generated (p) for the first 3 months of 2024 using historical wind data and forecasted independent variables.
- Create a database distributing the predicted total power (p) to 3 consumption nodes based on the 4-Node Star Architecture (20%, 45%, and 35%).
- Use SVM Classification to assess grid stability based on the prepared dataset, including:
- Predicted power generated and stored at each node
- Existing data from the "Grid folder" (price per unit, unit consumption, and grid stability reports)

ARIMA Model:

Strengths:

- Specifically designed for time series forecasting, which is ideal for predicting future wind power generation based on historical data.
- Can capture trends and seasonality present in the data.
- Often requires less data compared to Random Forest.

Weaknesses:

- Assumes a linear relationship between past observations and future values. This may not hold true if the wind patterns are highly non-linear.
- Limited ability to handle complex relationships between multiple features (Air Temperature, Pressure, Wind Speed).

Additional Notes:

- You can explore feature engineering techniques to improve model performance for both tasks.
- Consider metrics like Mean Squared Error for wind power generation prediction and accuracy/F1-score for grid stability assessment to evaluate your models.
- The hackathon encourages exploring beyond the suggested examples in the final report. Look for patterns in grid instability and suggest mitigation strategies based on your findings.

Tools and Libraries:

Data Manipulation: Pandas
Time Series Analysis: Pandas, Statsmodels
Machine Learning: scikit-learn
Visualization: Matplotlib, Seaborn

Implementation Phases:

1. Data Preprocessing and Exploration (Both Datasets):

- Load Data: Use Pandas to read wind energy and smart grid data (CSV files).
- Handle Missing Values: Impute missing values using techniques like mean/median imputation or forward fill for time series data.
- Feature Engineering (Wind Data):
  - ❖ Create new features like wind power density or wind speed squared.
  - ❖ Consider time-based features like seasonality or day of the week.
- Data Cleaning (Grid Data):
  - ❖ Ensure consistency in units and data types.
  - ❖ Identify and handle outliers if necessary.
  - ❖ Merge predicted power generation data with grid data.

2. Wind Power Generation Forecasting (Task 1):

- Model Selection: Implement Random Forest Regression.
- Train-Test Split: Split historical wind data (excluding 2024) into training and testing sets.
- Train Model: Train the Random Forest model on the training data with historical power generation as the target variable.
- Hyperparameter Tuning: Optimize model parameters using techniques like GridSearchCV to improve performance.
- Evaluation: Evaluate model performance on the test set using Mean Squared Error (MSE).
- Prediction: Use the trained model to predict total power generation (p) for the first 3 months of 2024 based on the provided forecasted independent variables.

3. Smart Grid Power Distribution (Task 3):

- Create Database: Based on predicted total power (p), create a new data structure (e.g., Pandas DataFrame) with columns for:
    - ❖ Date/Time
    - ❖ Predicted Power Generation (p)
    - ❖ Power Distributed to Node 1 (20% of p)
    - ❖ Power Distributed to Node 2 (45% of p)
    - ❖ Power Distributed to Node 3 (35% of p)

4. Smart Grid Stability Assessment (Task 4):

- Model Selection: Implement Support Vector Machine (SVM) Classification.
- Feature Selection: Choose relevant features from the combined dataset (predicted node power, existing grid data).
- Train-Test Split: Split the combined dataset into training and testing sets.
- Train Model: Train the SVM model on the training data with grid stability ("stable" or "unstable") as the target variable.
- Hyperparameter Tuning: Optimize SVM parameters for improved classification accuracy.
- Evaluation: Evaluate model performance on the test set using accuracy or F1-score.
- Prediction: Use the trained SVM model to predict grid stability for the first 3 months of 2024 based on the combined dataset.

5. Analysis and Reporting:

- Grid Stability Analysis:

    - ❖ Analyze the percentage of "stable" and "unstable" grid conditions over the 3 months.
    - ❖ Identify patterns in grid instability occurrence (time of day, specific weather conditions).

- ❖ Suggest mitigation strategies based on your findings (e.g., adjusting power generation or consumption during peak instability periods).
- ❖ Visualization: Create insightful visualizations (e.g., time series plots, heatmaps) to showcase results and trends.
- ❖ Reporting: Prepare a comprehensive report including:
- ❖ Methodology and model selection rationale
- ❖ Model performance metrics
- ❖ Grid stability analysis and insights
- ❖ Recommendations for grid operators

Additional Considerations:

- Explore feature engineering techniques for both wind and grid data to potentially improve model performance.
- Consider using time series forecasting techniques like ARIMA for wind power generation if you suspect a linear relationship in the data.
- Experiment with different machine learning models for grid stability assessment (e.g., Random Forest) and compare their performance.
- Explore more advanced analysis techniques like anomaly detection to identify unusual grid behavior.