**Research Plain**

**1. Key Hypotheses and Questions**

1. **Which kinds of errors is the most critical ( causing the biggest cost)?**

**Why?** Knowing which faults incur the highest costs allows us to prioritize preventive maintenance and fault detection strategies.

**How? Define the cost!**: Study the direct costs of the errors (repair cost, maintenance cost, labor) and indirect costs (downtime, lost production, default of the contract). Additionally, consider the long-term costs for the wind farm. (such as the impact on the lifespan of turbines)

1. **Are Inverter Conditions the Main Cause of FF Errors?**

**Why?** To pinpoint the main reason that leads to the most frequent turbine malfunction.

**How?** Build a model by selecting some key features (How to?), including inverter temperature, power output, error logs, and maintenance records to do a causal inference.

**2-1. Data Collection(data we have)(data needs to collect)**

**Error Logs**: Detailed records of turbine malfunctions, timestamps, and fault types.

**Maintenance Records**: Data on maintenance actions, parts replaced, and their timing.

**Operational Data**: Measurements of turbine performance, including inverter temperature, power output, and environmental factors.

**Cost Data**: Information on direct and indirect expenses related to repairs, downtime, and lost energy production.

**2-2. Essential Features(to validate hypothsis):**

**Must-have:**

* **Inverter Temperature**: Identifies overheating or performance degradation.
* **Signal of the Power Outage**: Monitors fluctuation that may signal inverter issues.
* **Detailed Error Logs and Timestamps**: Provides a history of failures to assess trends.
* **Maintenance Records**: Links maintenance actions to error occurrences.

**Nice-to-have:**

* **Other Potential Error Control Methods:** What other factors may lead to failure? Are there any chain effects?
* **Detailed Component-level data:** What specific parts should be replaced during maintenance?

**3. Analysis and Methodology**

**Cost Analysis**

* **Goal**: Identify which errors have the highest associated costs.
* **Approach**: Calculate average and variance of repair, downtime, and lost production costs per error type. Rank errors based on total cost impact.

**Correlation Analysis**

* **Goal**: Assess the relationship between inverter conditions and FF errors.
* **Approach**: Analyze correlations between inverter temperature, power output fluctuations, and occurrences of FF errors using regression and time-lagged analysis.

**Feature Selection**

* **Goal**: Determine predictive features for FF errors.
* **Approach**: Use Recursive Feature Elimination (RFE) with a machine learning model to identify top predictors among operational metrics.

**Hypothesis Testing**

* **Goal**: Test if inverter conditions are a primary cause of FF errors.
* **Approach**: Use logistic regression for binary outcome analysis (FF error vs. no error) or survival analysis to assess time-to-failure based on inverter conditions. Tree-based models, like random forests, may help capture non-linear relationships and variable importance.

**4. Key Findings**

* **Top Costly Errors**: What’s the most urgent error we must deal with?
* **Inverter and FF Error Correlation**: Can we mitigate the FF error by maintaining the inverter more often?
* **Predictive Features for FF Errors**: How good is our predictive ability for the turbine failure based on our selected inverter data, such as temperature, power output, and maintenance frequency?

**5. Prescriptive analysis**

1. **Improve the Implementation of Real-Time Monitoring**: Set thresholds for inverter temperature and power output fluctuations to trigger early warnings.
2. **Schedule Maintenance Based on Error Costs**: Focus preventive actions on high-cost faults identified through cost analysis.
3. **Enhance Predictive Maintenance Strategy**: Integrate predictive features into maintenance schedules to proactively address high-risk conditions.