Team Submission for the 2024 TUMai Hackathon

Challenge Sponsor: REPLY

Challenge: Develop predictive maintenance solutions to optimize energy production and prevent unexpected downtimes in wind turbines.

Introduction

The challenge posed by REPLY was to leverage wind turbine system data to develop predictive maintenance strategies that would enhance energy efficiency and operational reliability. Our goal was to minimize downtime, maximize power output, and extend the lifespan of wind turbines using machine learning techniques.

Approach and Methodology

Data Review and Selection

- Initial data provided by REPLY lacked failure signals; only weather data was available which was insufficient for predictive analysis.
- Sourced additional datasets from Penmanshiel and Kelmarsh Wind Farms in England, containing over 400 data points including failures and SCADA sensor data.
- Selected Kelmarsh Wind Farm data for initial focus due to its comprehensive data on failures and sensor readings.

Step-by-Step Process

1. **Understanding Failures**
   - Evaluated failures based on their frequency, predictability, and preventability.
   - Identified economic impacts and assessed if maintenance could prevent damages.
   - Focused on frequency converter errors due to time constraints, aiming to predict and prevent this specific failure.
2. **Feature Reduction**
   - Analyzed datasets to identify where less than 10% of the data were NaN values.
   - Performed technical reduction to determine 48 relevant features that would be crucial for the machine learning model.
3. **Error Data Analysis**
   - Quantified the occurrence of frequency converter errors and the associated downtime.
4. **Feature Data Analysis**
   - Conducted visual inspections comparing damage periods with random periods to detect patterns and correlations.
   - Repeated this analysis for different time frames (daily, weekly, multi-weekly) to ensure robustness.
5. **Visualization and Dimension Reduction with Umap**
   - Explored how error labels distributed across different conditions.

Friedrich Siemers, Phillip Wondra, Benedikt Kellner, Maximilian Seeger

- Detected clusters, although noted redundancy due to multiple counts of events per day.
6. **Data Cleaning**
   - Consolidated multiple error messages from the same day into a single record to refine the dataset.
7. **Insufficient Data for Reliable Predictions**
   - Post-cleaning, it was evident that the dataset contained limited failure data, with only 14 points available for predictive modeling.
8. **Time Series Analysis**
   - Developed time series data in intervals of 24, 48, and 72 hours before damage.
   - Created aggregated features such as mean and standard deviation within these windows to enhance model accuracy.
9. **Further Feature Reduction and Predictive Modeling**
   - Applied feature importance metrics and techniques like ANOVA and recursive feature elimination with logistic regression to minimize overfitting.
   - Generated a logistic regression model to predict future frequency converter errors.
10. **Model Testing**
   - Applied the developed model to the Penmanshiel dataset for validation, which did not yield successful predictions due to data limitations.

Challenges Encountered

- **Data Limitations:** The scant amount of failure data restricted the statistical relevance and the training robustness of our predictive model.
- **Short Data Span:** Limited error data points undermined the ability to train a model capable of generalizing across different operational scenarios.

Conclusion and Recommendations

Despite the initial setbacks due to data insufficiency, our team was able to develop a methodology to identify and predict critical failures in wind turbine operations, specifically targeting frequency converter errors. For future work, acquiring more comprehensive datasets with a broader range of failure instances is crucial to refining the predictive models. Our efforts have laid the groundwork for developing predictive maintenance strategies that can significantly reduce downtime and enhance the efficiency of wind turbines.

Future Directions

- Expand data collection efforts to include a wider array of error types and operational conditions.
- Explore more advanced machine learning algorithms that can handle sparse data more effectively.

Friedrich Siemers, Phillip Wondra, Benedikt Kellner, Maximilian Seeger