Title: Robust Power Prediction of Wind Turbine using Error Detection, Clustering-Based Imputation, and Physics-Informed Learning

Authors: Swayam Mittal, Vishwaas Narasinh, Nikhil Kulkarni, Remish Leonard Minz, Nilanjan Chakravortty, Prateek Mital

Affiliation: Hitachi India Ltd., Bangalore, India

Introduction

- Introduction to Wind Energy
  - Significance of wind energy in the renewable spectrum

    Wind energy plays a critical role in the renewable energy spectrum by providing a sustainable and eco-friendly solution to meet growing global energy demands. Its increasing deployment reduces reliance on fossil fuels, significantly contributing to the mitigation of climate change.

  - Challenges in power prediction due to erroneous and missing data

    Challenges in power prediction for wind turbines arise due to the presence of erroneous and missing data, which can lead to inaccuracies in the models. These data issues disrupt the reliability of power forecasts, complicating grid stability and efficient energy management.

Research Motivation

- Enhancing Power Prediction Accuracy
  - Need for reliable power predictions for grid stability and energy management

    Reliable power predictions are essential for grid stability and efficient energy management, as they help balance supply and demand, prevent blackouts, and optimize the integration of renewable energy sources into the power grid. Accurate forecasts ensure that wind energy contributions are effectively utilized, enhancing overall grid performance and sustainability.
  - Traditional methods vs. advanced techniques

    Traditional methods for wind turbine power prediction often rely on basic statistical models and simple imputation techniques, which can struggle to handle the complex, non-linear relationships in the data. Advanced techniques, such as auto-encoders and Physics-Informed Neural Networks (PINNs), leverage machine learning and physical principles to improve accuracy, effectively addressing the intricacies of erroneous and missing data.

Related Work

- Key Approaches and Limitations
  - Traditional anomaly detection and imputation methods

    Traditional anomaly detection methods use simple statistical techniques that may miss complex patterns in the data. Basic imputation methods like mean or median substitution often oversimplify data, leading to inaccuracies in power prediction models.
  - Recent advancements using auto-encoders and physics-informed models

    Recent advancements involve using auto-encoders for anomaly detection and data imputation, which better capture complex patterns in wind turbine data. Physics-Informed Neural Networks (PINNs) integrate physical laws with data-driven models, enhancing prediction accuracy and ensuring physically plausible results.

Data Collection Overview

- Dataset Characteristics

Characteristic	Description
Wind Farm Location	Gujarat, India
Operational Period	November 1st, 2022 - July 15th, 2023
Number of Turbines	16 wind turbines
Total Capacity	32 MW
SCADA Tags	1966 tags from various turbine components
Data Recording Frequency	Approximately every 2 minutes
Data Averaging Interval	10-minute intervals
Focus Turbines	2 wind turbines
Filtered Data Criteria	Wind speed < 3 m/s or > 10.5 m/s, and no production (0 kWh)

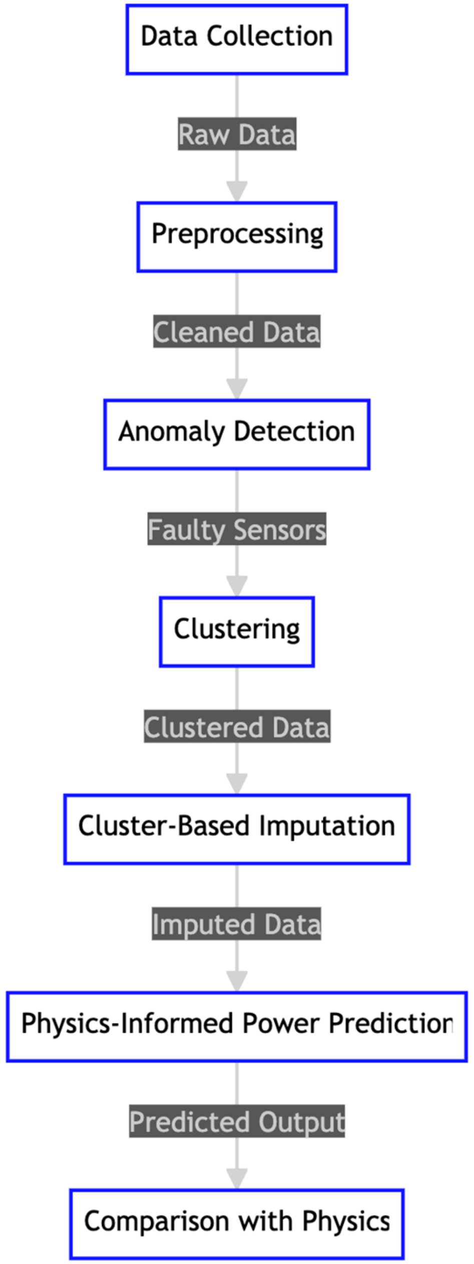
Handling Outliers

- Identification and Methodology
  - Types of outliers observed
    - **Non-operating Phase Outliers:** These occur during maintenance phases when the turbine's active power is zero despite wind speeds being within operational ranges. These data points were identified as non-operational and removed to avoid misinterpretation.
    - **Power Curve Deviation Outliers:** These data points significantly deviate from expected values on the power curve, potentially due to issues like dirt accumulation on blades, or pitch malfunctions. These deviations indicate operational anomalies that need to be addressed for accurate power prediction.

  - Steps for outlier detection and removal
    - Outlier detection involved visually examining wind speed versus power plots, using interval-based methods to flag deviations, and validating against the power curve to identify significant anomalies. These steps ensured the removal of non-operational phase and power curve deviation outliers.

Methodology - Overview

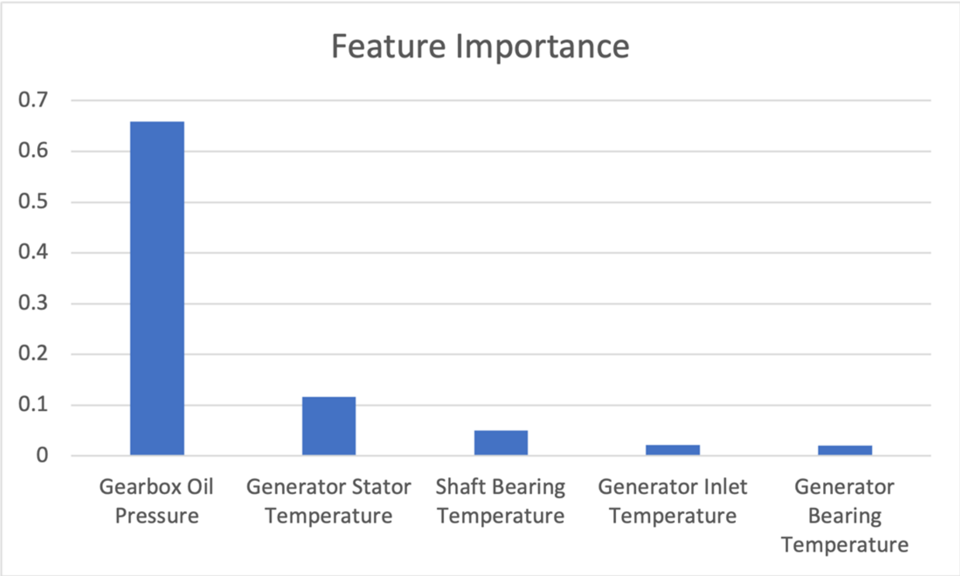
- Overview of the data processing and analysis phases
  The data processing and analysis phases involve detecting and handling outliers, selecting key features, performing anomaly detection, and employing clustering-based imputati before integrating Physics-Informed Neural Networks (PINNs for robust power prediction. This comprehensive pipeline ensures accurate and reliable wind turbine power forecasts.



Feature Selection

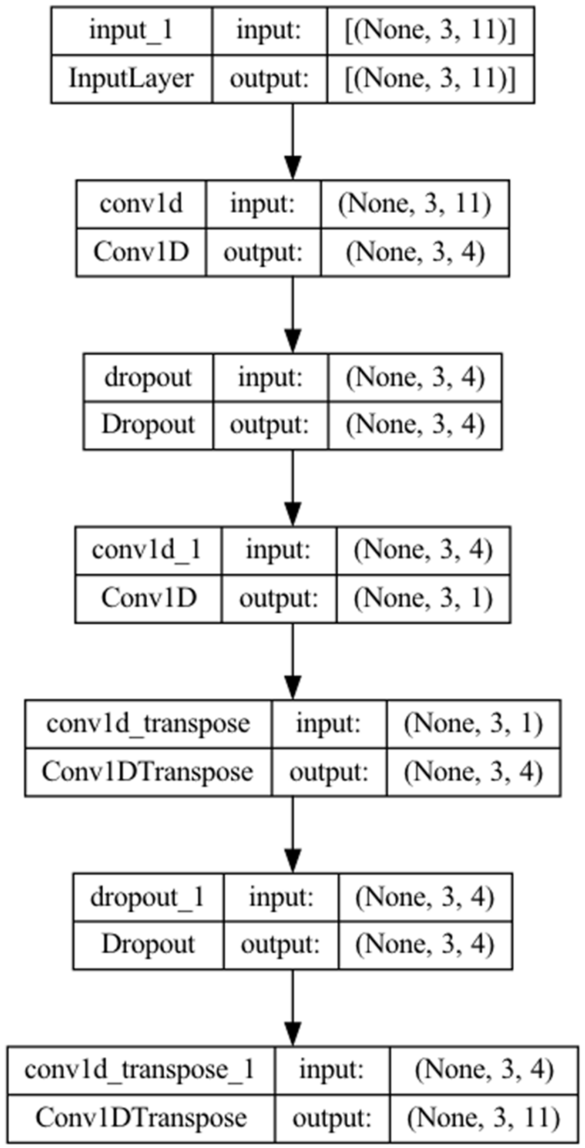
- Importance and Selection Process
  - Initial selection guided by domain knowledge and data availability

    The initial feature selection was guided by domain knowledge and data availability, ensuring the inclusion of relevant SCADA parameters crucial for accurate power prediction. Key features were shortlisted based on their significance to turbine operations and their contribution to the prediction model's performance.
  - Random Forest model for feature importance



Anomaly Detection

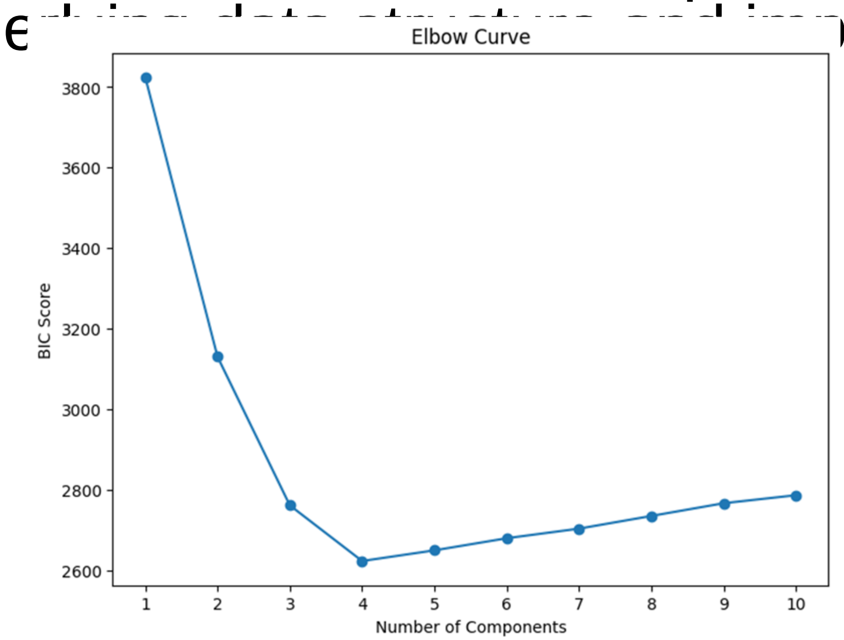
- Auto-encoder Based Approach for the Detection of faulty sensors
  - The auto-encoder architecture for anomaly detection consists of an encoder that compresses the input data into a latent representation and a decoder that reconstructs the input. The model is trained using normal operational data to minimize the reconstruction error. Faulty sensors are detected by identifying instances where the reconstruction error exceeds a predefined threshold, indicating anomalies in the sensor readings.



Clustering-Based Imputation

- Gaussian Mixture Models (GMM) for Imputation of missing values

  Gaussian Mixture Models (GMM) were used to cluster turbines based on multiple features, capturing the complex distributions of the data. This clustering enabled effective imputation of missing values by replacing them with the mean values from the corresponding clusters, thereby preserving the underlying data structure and improving prediction accuracy.



Power Output Modeling

- Neural Network Architecture with Integration of physics-informed loss

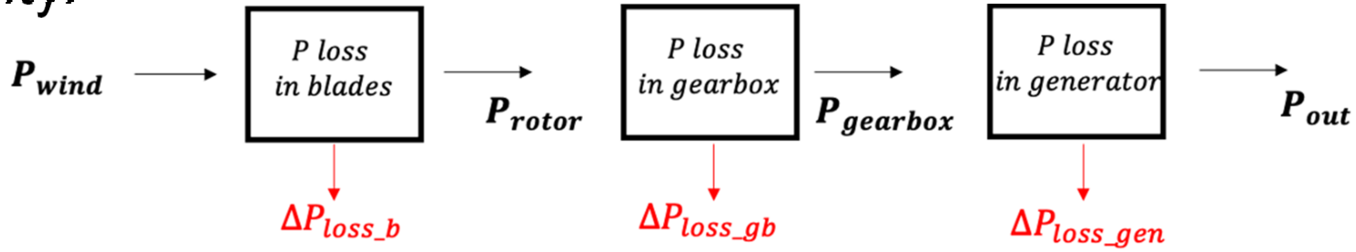
  Description of neural network layers and training process : The power output model uses a neural network with multiple fully connected layers, each employing ReLU activation functions and dropout for regularization. The network is trained using a 3-fold cross-validation approach, optimizing the Mean Absolute Error (MAE) loss. Additionally, a physics-informed loss function, derived from energy conservation laws governing wind turbines, is integrated into the training process to ensure predictions are both accurate and physically plausible.

  - [Patent NN diagram for power loss modelling with PINN]

Physics-Informed Loss

- Energy Conservation Laws
  - Derivation of physics-based loss function

    The physics-informed loss function is derived from energy conservation laws, accounting for the power in the wind, losses at the blades, gearbox, and generator. This loss function is applied to the power prediction model to ensure that the predicted output adheres to physical principles, thereby improving the model's accuracy and reliability.



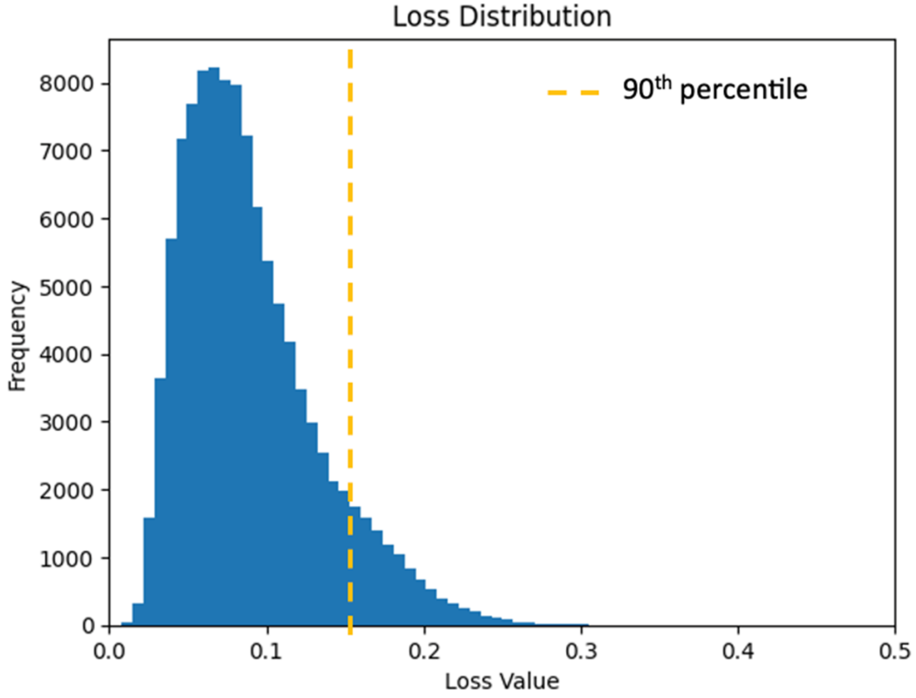
Imputation Accuracy

- Imputation Accuracy Comparison
  - Comparison of GMM, Autoencoder, K-means, and Simple Average methods

Method	**MAE (Imputation Accuracy)**
GMM	9.21
Autoencoder	12.62
KMeans	15.86
Simple Average	32.69

Validation of Anomalies

- Sensitivity Analysis
  - The sensitivity analysis involved testing the model by introducing various anomaly scenarios, such as altering sensor values to simulate faulty readings. The model's response to these scenarios was evaluated by measuring the reconstruction error, confirming its ability to detect anomalies accurately and ensuring the robustness of the anomaly detection mechanism.



Evaluation Metrics

- Performance Metrics
  - Accuracy of power prediction
    - Measure how closely the predicted power values match the actual values using the coefficient of determination (R^2).
    - Evaluate the improvement in prediction accuracy with the integration of physics-informed loss.
  - Robustness of imputation techniques
    - Assess the effectiveness of various imputation methods in handling missing and erroneous data.
    - Compare the performance of GMM-based imputation against other techniques like autoencoders and simple average imputation.

  - [Table comparing R^2 values with and without physics loss]

Faulty sensor	**Without Physics Loss**	**With Physics Loss**
Wind Speed	0.67	0.77
Rotor Speed	0.51	0.58
None	0.77	0.79

Ablation Study

- Feature Contribution
  - Impact of each feature on prediction accuracy
    - Conduct an ablation study by individually adjusting each feature from its mean value.
    - Compare the prediction accuracy with and without the physics-informed model to evaluate the significance of each feature.

  - [Table showing accuracy with and without physics model for different feature deviations]

Feature	**Deviation**	**Acc w/o Phy**	**Acc w/ Phy**
Gearbox Oil Pressure	+/- 1%	93.2%	94.1%
Gen. Starter Temp.	+/- 0.7%	92.5%	93.8%
Shaft Bearing Temp.	+/- 0.6%	93.0%	93.5%
Gen. Inlet Temp.	+/- 0.7%	92.8%	93.7%
Gen. Bearing Temp.	+/- 0.6%	93.1%	93.9%
Pitch Angle	+/- 0.5%	92.4%	93.3%
Wind Speed	+/- 2%	91.9%	93.0%
Rotor Speed	+/- 1%	93.2%	94.2%
Nacelle Direction	+/- 0.8%	92.6%	93.6%
Yaw	+/- 0.5%	92.3%	93.2%

Results and Discussion

- Key Findings
  - Improvement in prediction accuracy with physics-informed loss
    - Incorporating physics-informed loss into the power prediction model significantly enhances accuracy.
    - The R^2 values show marked improvement across different scenarios, demonstrating the effectiveness of integrating physical principles with data-driven models.
  - Better imputation accuracy using GMM
    - The Gaussian Mixture Model (GMM)-based imputation method outperforms traditional techniques such as K-means and simple average imputation.
    - GMM effectively handles the complex, non-linear associations in the data, resulting in lower Mean Absolute Error (MAE) and more reliable imputation.

Conclusion

- Summary of Contributions
  - Enhanced power prediction accuracy
  - Integration of data-driven and physics-based models

Future Work

- Potential Enhancements
  - Optimization of clustering techniques
  - Integration with real-time data

Thank -
You