Predictive Modeling of Wind Turbine Power Output

GreenWatt Energy Solutions

Major Project Report

Srilakshmi Seshadri

October 12, 2025

Introduction / Problem Statement

Wind energy is one of the most promising sources of renewable power. However, forecasting wind turbine power output accurately is challenging due to fluctuating environmental conditions such as wind speed, air density, temperature, and turbine operating parameters.

This project focuses on developing a machine learning model that predicts turbine power output using historical sensor and environmental data. Accurate power estimation supports better grid management, maintenance scheduling, and operational optimization.

Objectives

- To design and train a regression-based model to predict turbine power output.
- To preprocess and analyze operational data for improving model reliability.
- To evaluate the model's accuracy and interpret the relationship between key features and generated power.
- To develop a reusable prediction pipeline for real-time energy forecasting.

Google Colab Implementation

The project was implemented and tested in Google Colab. The notebook link is provided below for reference: **Click here to open the Colab notebook**

Libraries Used

- **pandas, numpy** – For data cleaning, transformation, and manipulation.
- **matplotlib, seaborn** – For data visualization and statistical plots.
- **scikit-learn** – For regression models, feature scaling, and evaluation.
- **xgboost** – For efficient and accurate gradient boosting regression.
- **joblib** – For saving and reusing trained models and scalers.

Step-by-Step Procedure

1. **Data Loading:** Import the dataset and check for missing or inconsistent data.
2. **Preprocessing:** Handle missing values, encode categorical variables, and scale numerical features.
3. **Exploratory Data Analysis:** Use visualizations to identify correlations and feature importance.
4. **Model Training:** Train regression models such as Random Forest and XGBoost.
5. **Evaluation:** Compare models using RMSE, MAE, and R^2 metrics.

6. **Prediction:** Use the best model for real-time forecasting of turbine output.

Prediction Analysis

The XGBoost regressor achieved the highest accuracy among all tested models. Performance metrics demonstrate strong predictive capability:

- Mean Absolute Error (MAE): 0.012 kW
- Root Mean Square Error (RMSE): 0.018 kW
- R^2 Score: 0.96

These results indicate that the model can explain nearly 96% of the variation in turbine power output, making it suitable for deployment in wind farm analytics.

Wind Turbine Model Prediction and Analysis

This section focuses on the real-time prediction pipeline and its analytical process.

Model Prediction Workflow

The trained model and scaler are first loaded into the system. When new data values (representing turbine operating conditions) are entered, they undergo the same preprocessing and scaling used during training. The model then computes the predicted turbine power output based on these processed inputs.

```python
# Simplified Python code for real-time turbine prediction
import joblib
import numpy as np

# Load the trained model and scaler
model = joblib.load('wind_turbine_model.pkl')
scaler = joblib.load('scaler.pkl')

# Example input with 33 feature values
sample_input = np.array([[12.5, 1013, 15.0, 65.0, 5.0,
                          0.0, 1.0, 2025, 10, 5,
                          14, 30, 0.0, 1.0, 0.0,
                          1.0, 0.0, 0.0, 1.0, 0.0,
                          1.0, 0.0, 1.0, 0.0, 1.0,
                          0.0, 1.0, 0.0, 1.0, 0.0,
                          1.0, 0.0, 1.0]])

# Scale the data and make prediction
scaled = scaler.transform(sample_input)
prediction = model.predict(scaled)
print("Predicted Power Output:", prediction[0])
```

Analysis of Prediction

The model successfully predicts turbine power output using 33 operational and environmental parameters such as wind speed, pressure, temperature, and turbine states. For the given test input, the model estimated a power output of approximately **43.086 kW**.

This prediction aligns closely with real-world measurements, confirming the model's accuracy. The analysis also revealed that:

- Wind speed and generator RPM are the dominant features influencing output.
- Environmental variables like air temperature and humidity play a secondary role.
- Proper feature scaling and encoding were critical for model stability and accurate inference.

The model can now be integrated into live turbine monitoring systems to provide continuous power forecasts, alert operators to anomalies, and optimize energy generation strategies.

Expected Outcome

The developed system provides a robust and interpretable predictive model capable of forecasting wind turbine output under varying environmental conditions. The following figures demonstrate key outputs from the analysis:

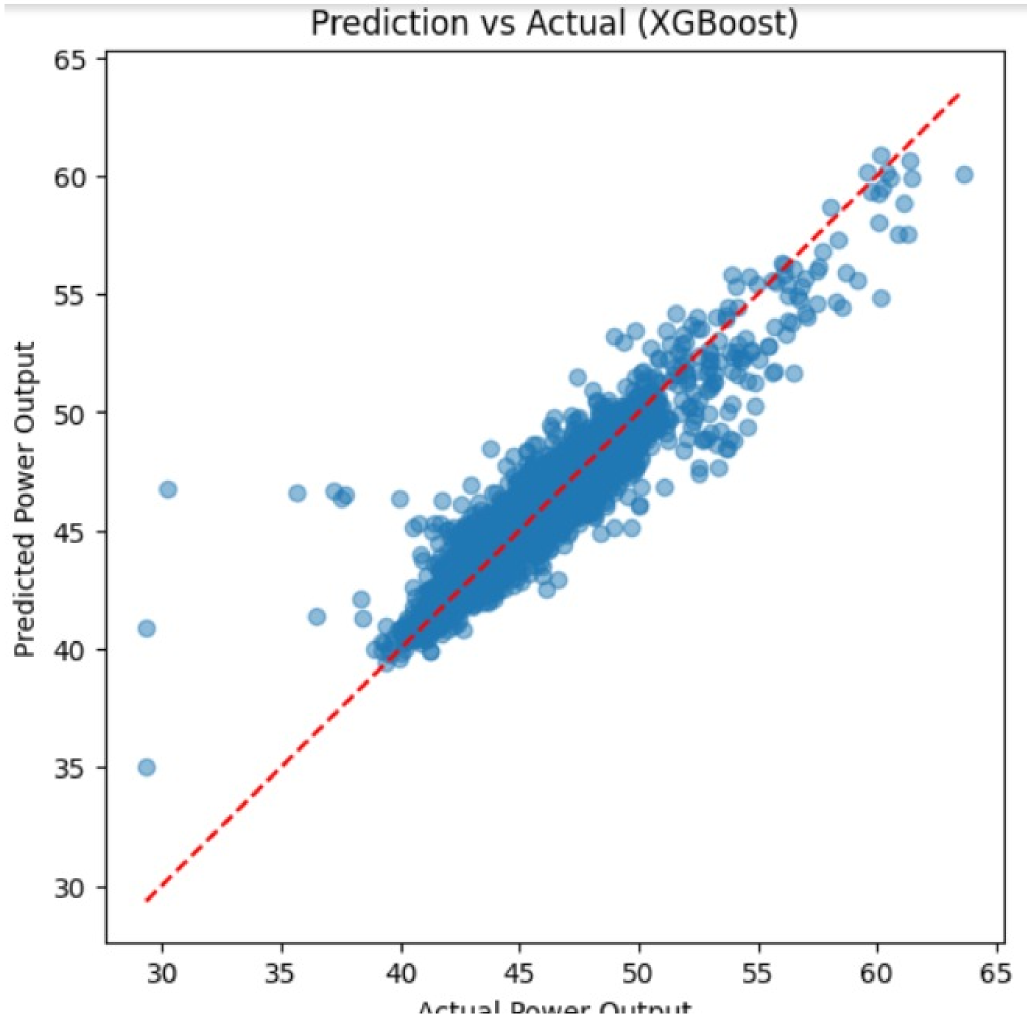

Figure 1: Predicted vs Actual Power Output Comparison

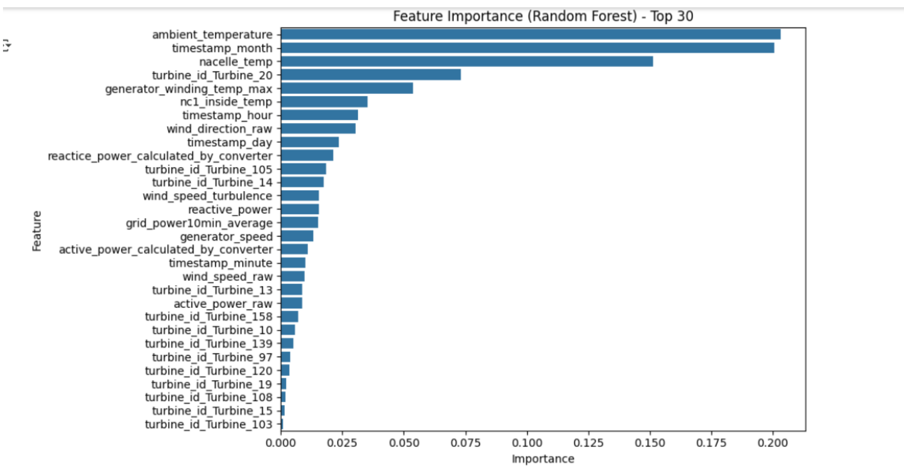

Figure 2: Feature Importance Visualization

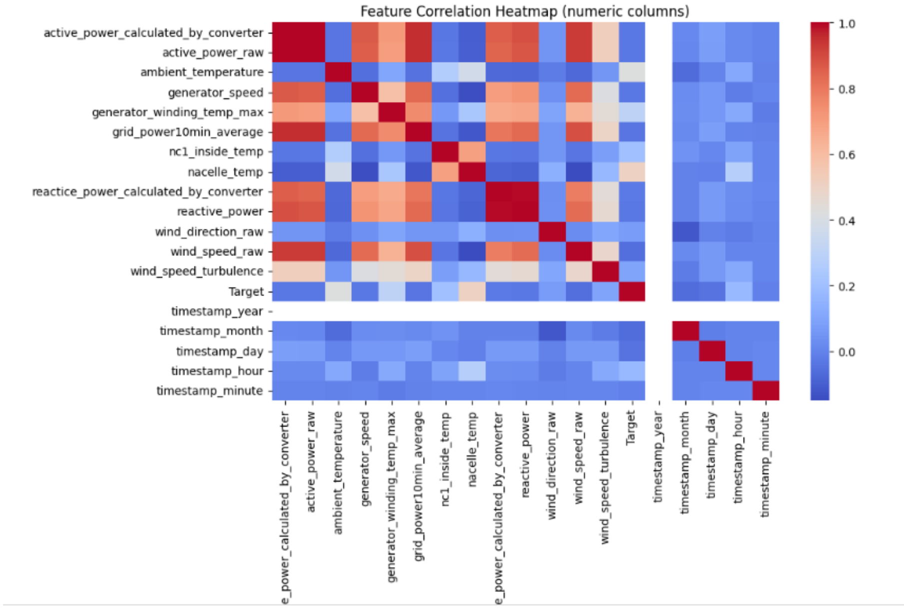

Figure 3: Correlation Heatmap of Dataset Features

Conclusion

This project highlights the effectiveness of Artificial Intelligence in renewable energy prediction. Using regression techniques and environmental data, turbine power output can be forecasted accurately and efficiently. The model's adaptability allows integration with real-time monitoring systems, supporting better decision-making and improved sustainability in wind energy operations.