Forecasting Wind Turbine Power Output Using Machine Learning Techniques

Wilson Hu
RISE NJ - Choice Research Group

Introduction

Figure 1: Distribution of energy sources for energy consumption in the United States (EIA, 2019)

Methodology

Phase 1: Dataset Preparation

A dataset from a wind turbine collected over a two-year period at 16-minute intervals (Bhaskarpandit, 2020) was prepared for model generation. Data entries that did not include date-time, wind speed, wind direction, active power, and reactive power were removed.

Phase 2: Model Generation

Linear Regression Model (random)
Random Forest Regressor
Multilayer Perceptron
Extreme Gradient Boosting Model
Support Vector Machine

Figure 3: Diagram of the main classifications of machine learning (IBM, 2020)

Wind speed (m/s)
Wind direction (°)
Ambient Temperature (°C)
Various turbine conditions (not incorporated)
All features selected

Learning curves for each model were generated to identify potential overfitting underfitting, and feature importances were plotted using XGBoost. K-fold cross-validation scores (k=10) were calculated to identify optimal models that minimize error.

Figure 4: Diagram of k-fold cross-validation (Packt, 2017)

Optimal Model

Cluster 1
Cluster 2
Cluster 3
Cluster n

Phase 3: Unsupervised Learning

Elbow test conducted to determine optimal number of clusters (n) followed by clustering using time-series k-means

Figure 5: Sample code block of model building using linear regression (figure by author)

Phase 4: Seasonal Analysis

Manual clustering based on season

Spring
Summer
Fall
Winter

Figure 6: Calculated average error values using k-fold cross-validation across model types (graph by author)

Wind speed (m/s)
Wind direction (°)
Ambient Temperature (°C)
Various turbine conditions (not incorporated)
All features selected

K-fold cross-validation scores (k=10) were calculated to compare seasonal clustering to unsupervised learning and supervised learning.

Phase 5: Hybrid Model

Combined two most optimal models to generate hybrid model

Supervised learning with all features
Unsupervised learning with all features
Seasonal Clustering with all features

Figure 7: Calculated average error values using k-fold cross-validation across learning types (graph by author)

K-fold cross-validation scores (k=10) were calculated to compare hybrid model to separated models.

Graphs were created in Microsoft Excel.

Results

Figure 8: Calculated average error values using k-fold cross-validation across feature selections (graph by author)

Figure 9: Feature importances plotted using XGBoost (graph by author)