Forecasting Wind Turbine Power Output Using Machine Learning Techniques

Wilson Hu
RISE NJ - Choice Research Group

Introduction

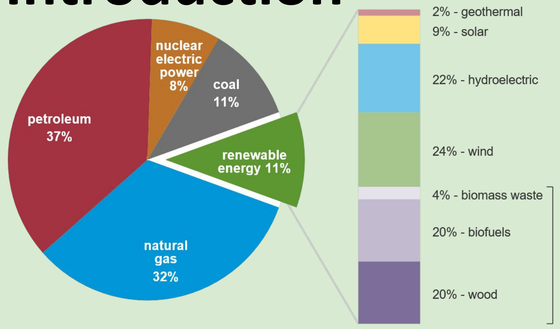

Figure 1: Distribution of energy sources for energy consumption in the United States (EIA, 2019)

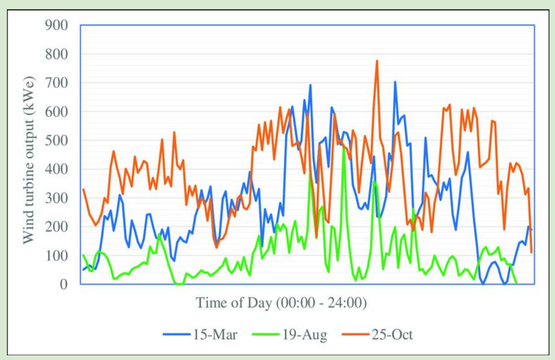

Figure 2: Daily power output of a wind turbine in Ireland (Huang et al., 2018)

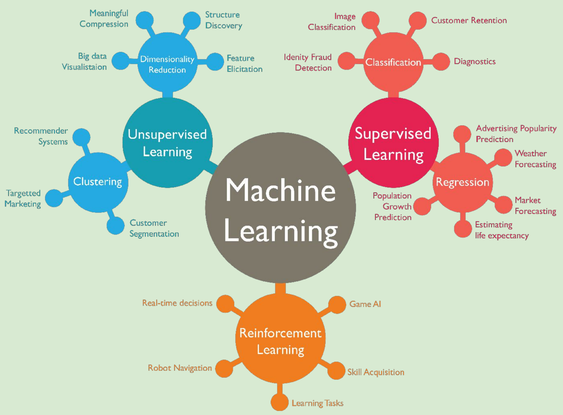

Figure 3: Diagram of the main classifications of machine learning (IBM, 2020)

Methodology

Phase 1: Dataset Preparation

A dataset from a wind turbine collected over a two-year period at 10-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 (control) | Random Forest Regressor | Multilayer Perceptron | Extreme Gradient Boosting Model | Support Vector Machine

Wind speed (m/s) | Wind direction (°) | Ambient Temperature (°C) | Various turbine conditions (shaft temperature/blade | 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.

Phase 3: Unsupervised Learning

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

Optimal Model

Cluster 1 | Cluster 2 | Cluster 3 | • • • | Cluster n

Wind speed | Wind direction | Ambient Temperature | Various turbine conditions | All features selected

K-fold cross validation scores (k=10) were calculated to compare models generated with and without unsupervised learning.

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

Phase 4: Seasonal Analysis

Manual clustering based on season

Spring | Summer | Fall | Winter

Wind speed (m/s) | Wind direction (°) | Ambient Temperature (°C) | Various turbine conditions (shaft temperature/blade angle) | 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

Hybrid model

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

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

Graphs were created in Microsoft Excel.

```
[ ] #linear regression
    model_1 = LinearRegression()

    #y_pred_model#_featureselection#
    #wind speed
    model_1.fit(X_1_train, y_1_train)
    y_pred_1_1 = model_1.predict(X_1_test)

    #wind direction
    model_1.fit(X_2_train, y_2_train)
    y_pred_1_2 = model_1.predict(X_2_test)

    #ambient temperature
    model_1.fit(X_3_train, y_3_train)
    y_pred_1_3 = model_1.predict(X_3_test)

    #turbine conditions
    model_1.fit(X_4_train, y_4_train)
    y_pred_1_4 = model_1.predict(X_4_test)

    #all four
    model_1.fit(X_5_train, y_5_train)
    y_pred_1_5 = model_1.predict(X_5_test)

    #cross-validation scores
    scores_1_1 = cross_val_score(model_1, X_1, y_1, cv=10, scoring='neg_mean_squared_error')
    scores_1_2 = cross_val_score(model_1, X_2, y_2, cv=10, scoring='neg_mean_squared_error')
    scores_1_3 = cross_val_score(model_1, X_3, y_3, cv=10, scoring='neg_mean_squared_error')
    scores_1_4 = cross_val_score(model_1, X_4, y_4, cv=10, scoring='neg_mean_squared_error')
    scores_1_5 = cross_val_score(model_1, X_5, y_5, cv=10, scoring='neg_mean_squared_error')

    print("Model 1, FeatSel 1:", scores_1_1.mean())
    print("Model 1, FeatSel 2:", scores_1_2.mean())
    print("Model 1, FeatSel 3:", scores_1_3.mean())
    print("Model 1, FeatSel 4:", scores_1_4.mean())
    print("Model 1, FeatSel 5:", scores_1_5.mean())

    Model 1, FeatSel 1: -46480.662037144865
    Model 1, FeatSel 2: -456149.9891995989
    Model 1, FeatSel 3: -439886.2044840255
    Model 1, FeatSel 4: -275456.5507457553
    Model 1, FeatSel 5: -55522.59837705372
```

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

Results

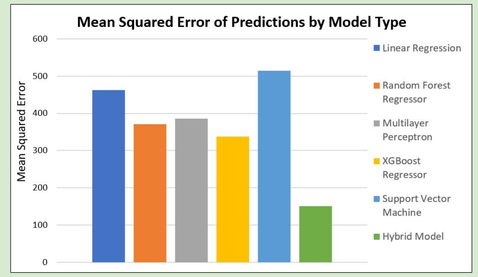

Figure 6: Calculated average error values using k-fold cross-validation across model types (graph by author)

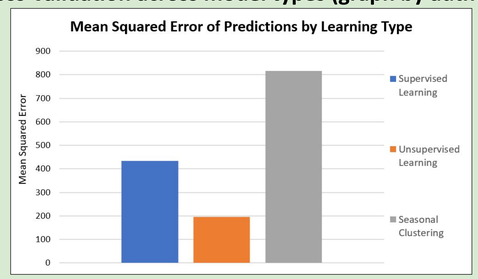

Figure 7: Calculated average error values using k-fold cross-validation across learning types (graph by author)

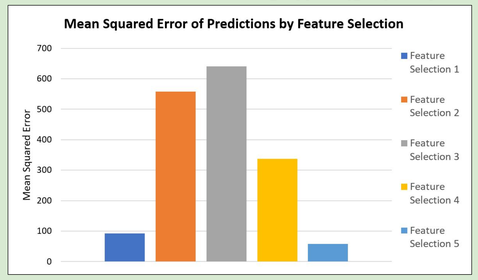

Figure 8: Calculated average error values using k-fold cross-validation across feature selections (graph by author)

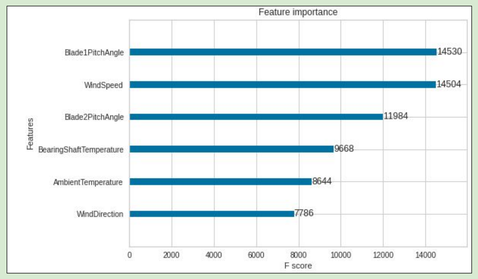

Figure 9: Feature importances plotted using XGBoost (graph by author)