ENGINEERING
TEXAS A&M UNIVERSITY

Boosting technique for power curve prediction

ISEN 619 Course Project

Jay Shah (826008661)
Bhaskar Botcha(626009084)

ENGINEERING
TEXAS A&M UNIVERSITY

Objective

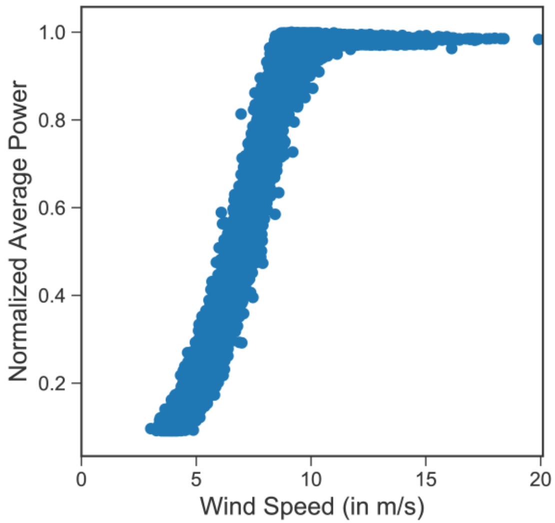
- To develop an effective learning method for power curve estimation with data collected from various sensors mounted on a wind mast from a wind farm.
- Apply the concepts learnt in class and explore new methods which can capture non-linear relationships seen in real-life sensor data.
- Obtain a test error (RMSE) lower than the industry ‘standard’, the IEC binning method.

Introduction

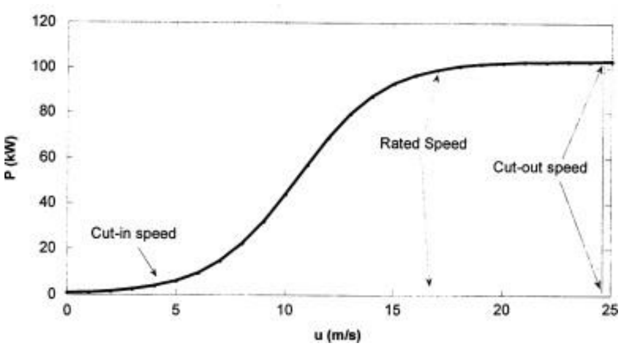
- According to DOE, wind energy will power up to 35% of the national demand by 2050, presenting a huge market potential.
- **Challenges:**
  - Power generation from a wind farm is highly unpredictable which is quintessential to monitor efficiency and health of turbines
  - Part of the uncertainty comes from the dearth of models which can predict wind power generation under dynamic environmental conditions.

Data visualization

- The power curve from data provided shows significant variability in it's operating region
- The variability cannot be captured with conventional ML models.
  - For example, to capture the high variability in using a linear regression, we impose heavy bias on the system by including many interaction terms which also results in over-fitting.
- The variability also suggests that wind speed not the sole contributor to predict the power output, but is affected by a plethora of other environmental factors which are not taken into account.



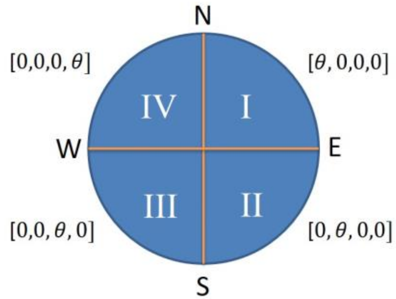
Power vs wind speed



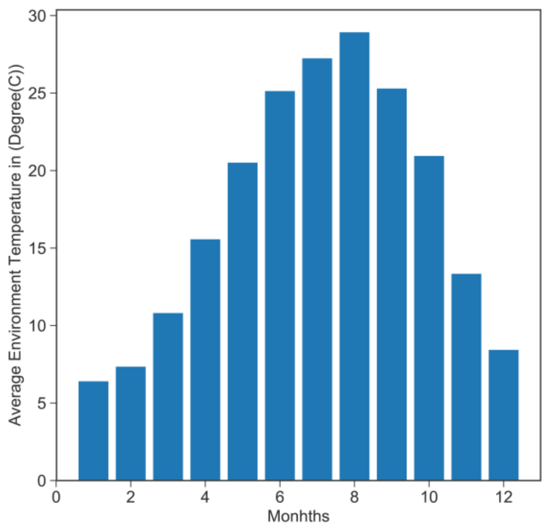
Nominal power curve

Feature engineering

- A few features like turbulence intensity I and air density ρ are derived from the existing feature set.
- Also, features for month(Mo), meridian(M) and direction vector (D_V) were used in the present implementation.
  - As air density is affected by temperature, and month and time of the day have significant effect on temperature, these variables were added to the feature set.
  - Also, the direction was converted into a vector to effectively capture the direction of wind velocity (improved result in CV)



Direction encoding

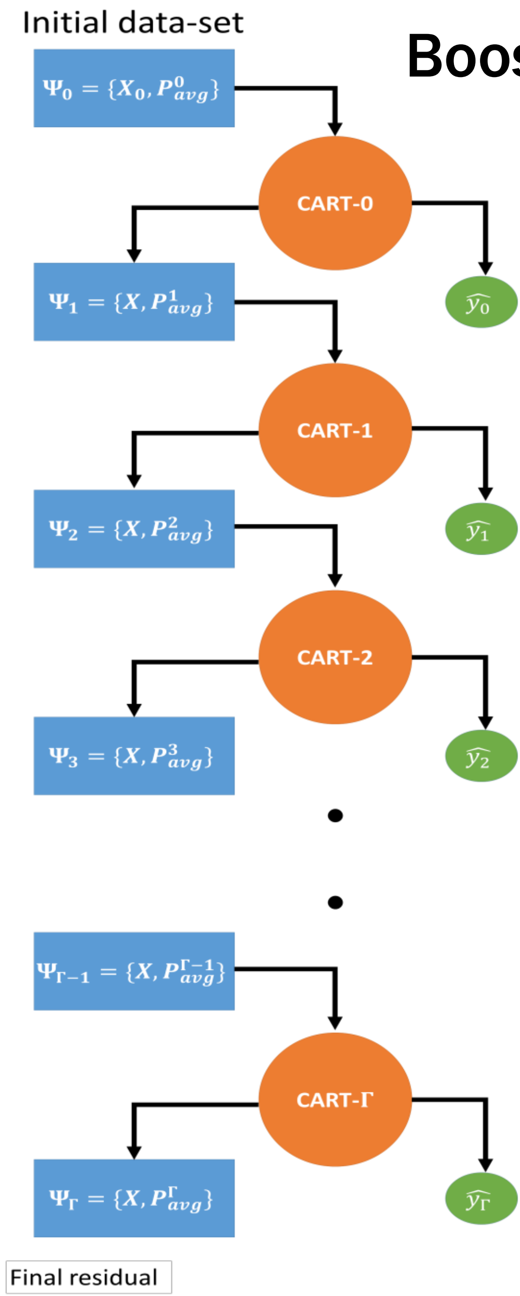


Temperature vs month

Boosting model

- For the present implementation, a gradient boosting method was used.
- Gradient boosting combines the weighted outputs of a set of base models to produce the final outcome. (Flowchart in next slide)
- <u>**Choice of boosting:**</u>
  - As capturing the high variability in the operating region of the turbine is fairly complex, we try fitting a series of models, where the residuals from the previous model are used to train the consequent models, thereby trying to capture the complex non-linear relationship in the data.

Boosting model, contd.



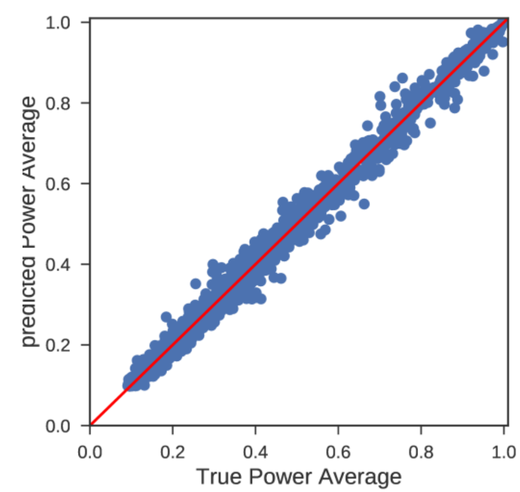
$$P_{avg}^1 = P_{avg}^0 - \widehat{y_0}$$

P_{avg}^1 are the residual values after fitting CART-0

$$\hat{y} = \widehat{y_0} + \widehat{y_1} + \ldots\ldots + \widehat{y_\Gamma}$$

Final prediction value $\hat{y}$ is the sum of all predictions after each CART. Each CART has a depth d, where each leaf node is assigned a value q_i. The sum of q_is gives y_is and the sum of y_is gives the total predicted output.

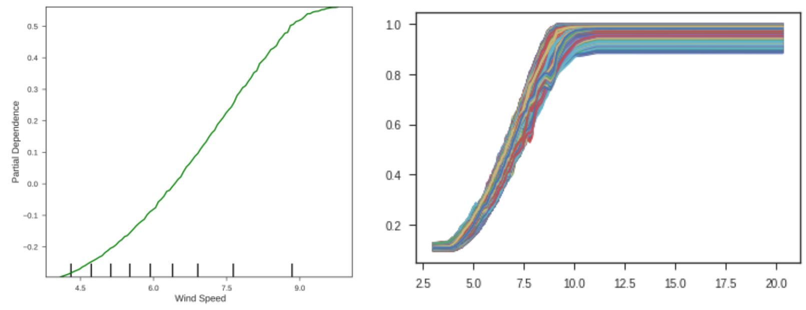
Parameters	Values
Number of trees	1000
Tree depth	8
Learning rate	0.01



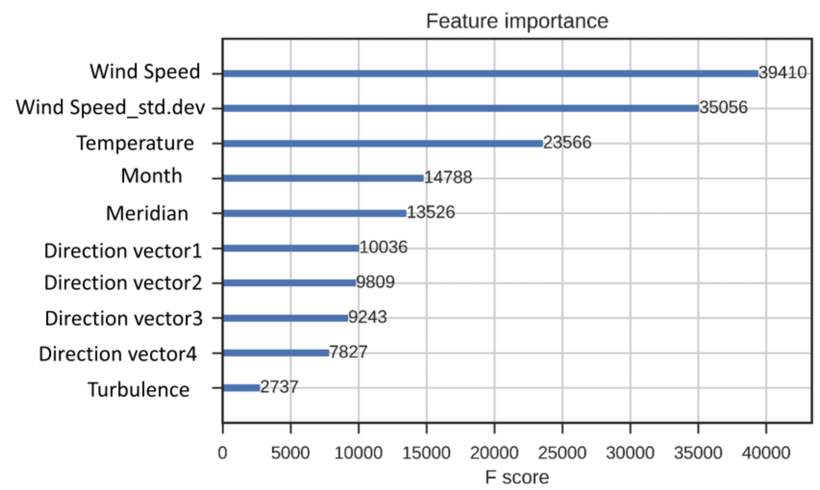
Predicted vs True values - XGBoost

Boosting model, contd.

- Black-box models like boosting can be analyzed through techniques like Partial Dependency Plots (PDPs) and Individual Conditional Expectation (ICE) plots.
  - <u>PDP</u> plots the change in the average predicted value as specific feature(s) vary over their marginal distribution.
  - <u>ICE</u> plots the estimated conditional expectation curves for each training data, where each curve represents the predicted response as a function of a co-variate, conditional on an observation. (In general, average of all ICE plots gives a PDP plot).



PDP and ICE plots for wind speed

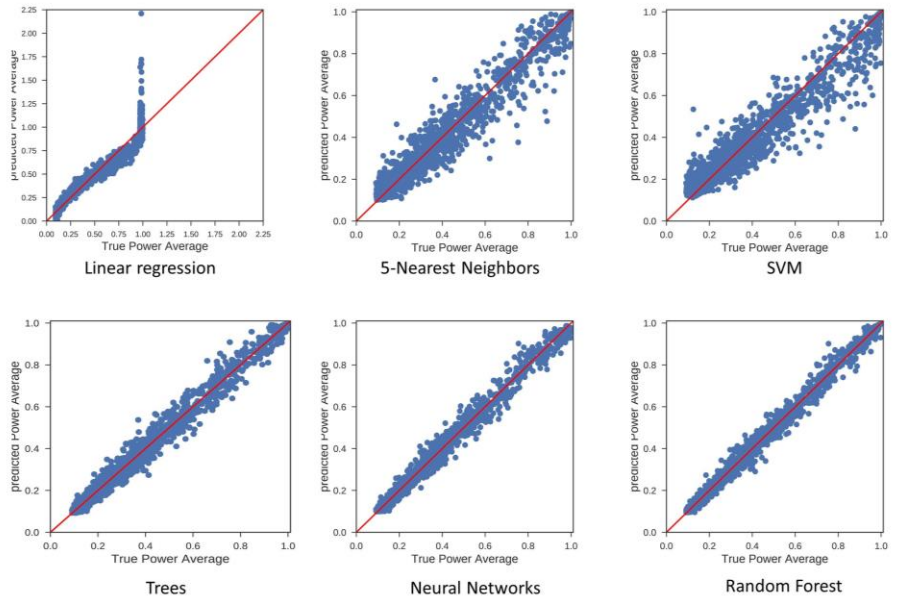


Feature importance plot

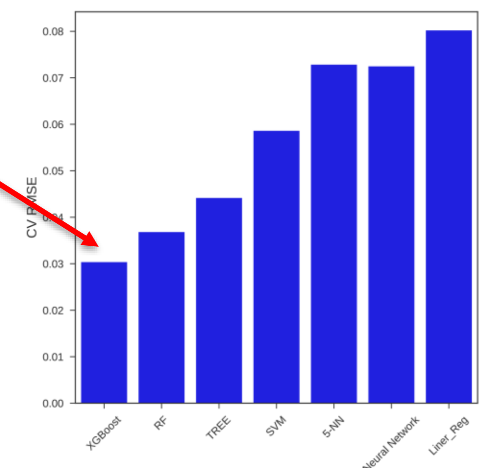
ENGINEERING
TEXAS A&M UNIVERSITY

Comparative analysis

- Analysis of different models shows that linear regression has the highest RMSE values.
- As cross-validated RMSE was the decision factor in deciding the best model, Boosting using XGBoost was finalized as the best model with RMSE=0.03, compared to the benchmark RMSE of 0.045 obtained through IEC binning



Final model
XGBoost



Conclusion

- In the current implementation, a boosting method was used for power curve prediction with least CV-RMSE and out-performing the industrial standard set by 'IEC binning'.
- The usage of boosting for power curve prediction was not found in literature, but was found to be better in contrast to other popular learning methods.
- Feature engineering was performed and cross-validated results for addition of new variables showed promising results, depicting that wind speed is not the only significant contributor for prediction of the power curve.
- With additional sensor data, few more variables like wind shear can be computed as found in literature, which can help capture the non-linearities better.