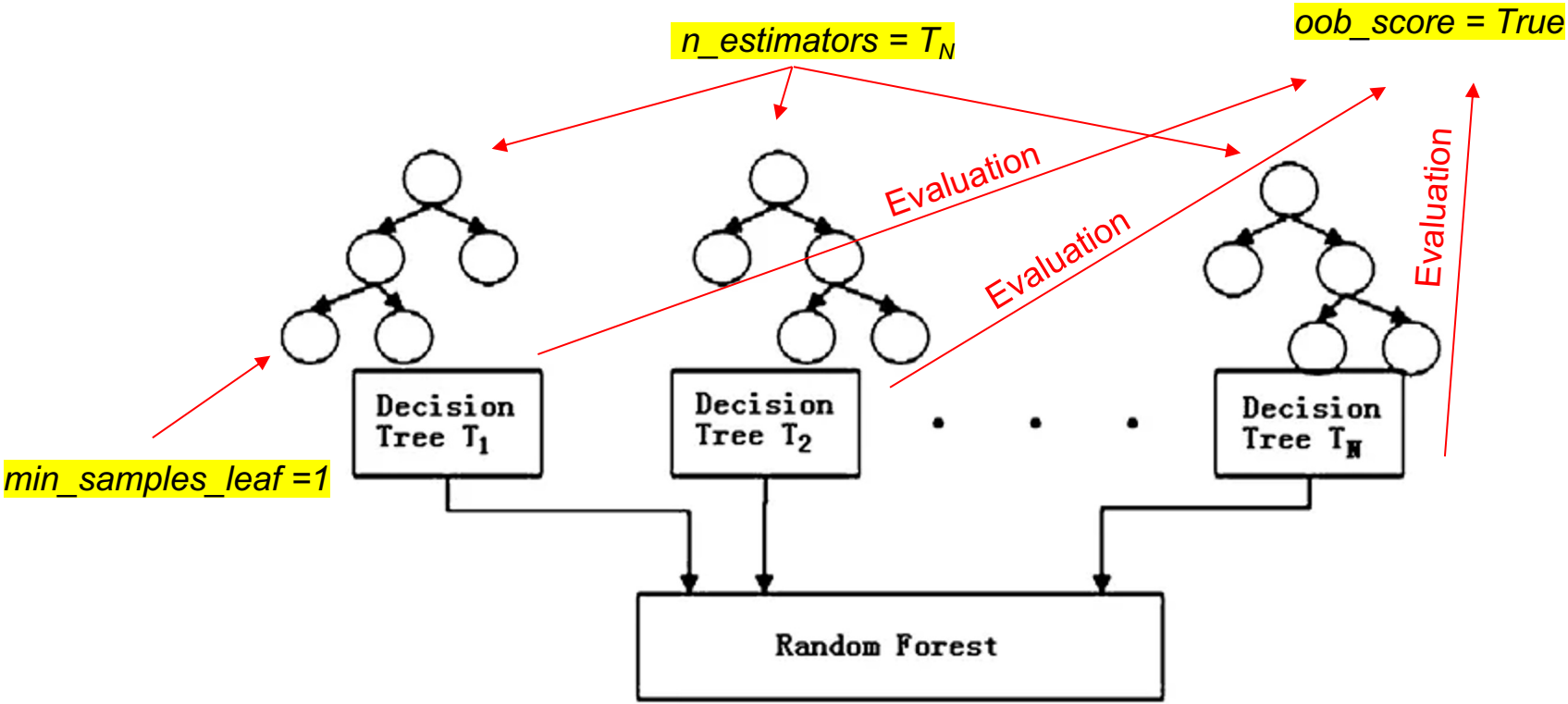
DATA-DRIVEN TURBINE PERFORMANCE ANALYSIS

Team 7
Texas A&M University

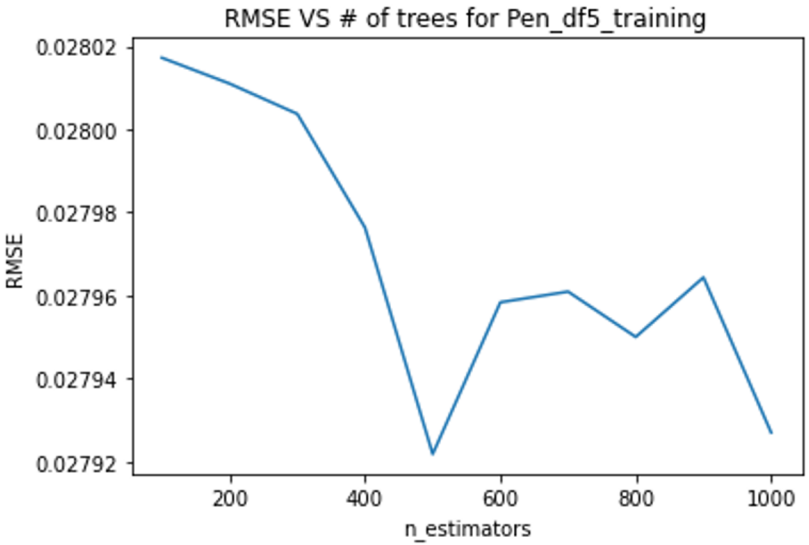
Random Forest: Simple Model

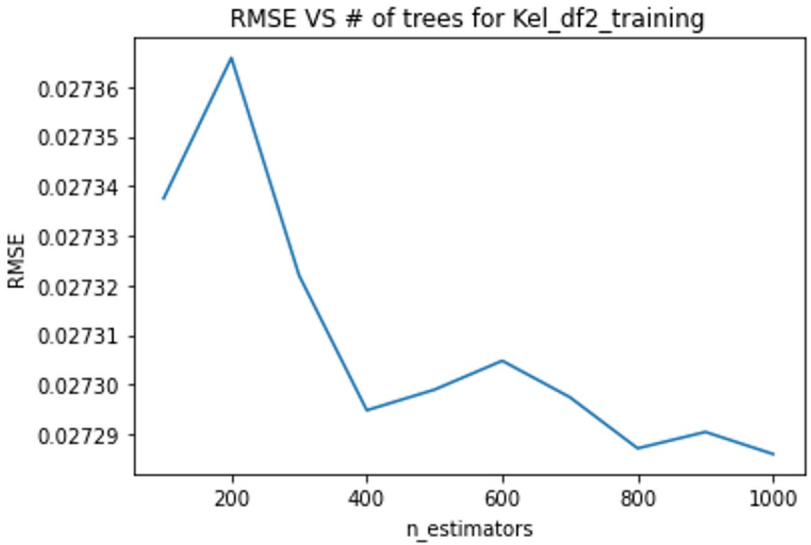


https://medium.com/all-things-ai/in-depth-parameter-tuning-for-random-forest-d67bb7e920d

Random Forest: Training & Tuning

- Split 80% training and 20% for validation.
- Train on one turbine per location: not optimal, but faster.





Random Forest: Results & Remarks

Criteria	MAE	RMSE
Score	0.0160	0.0281

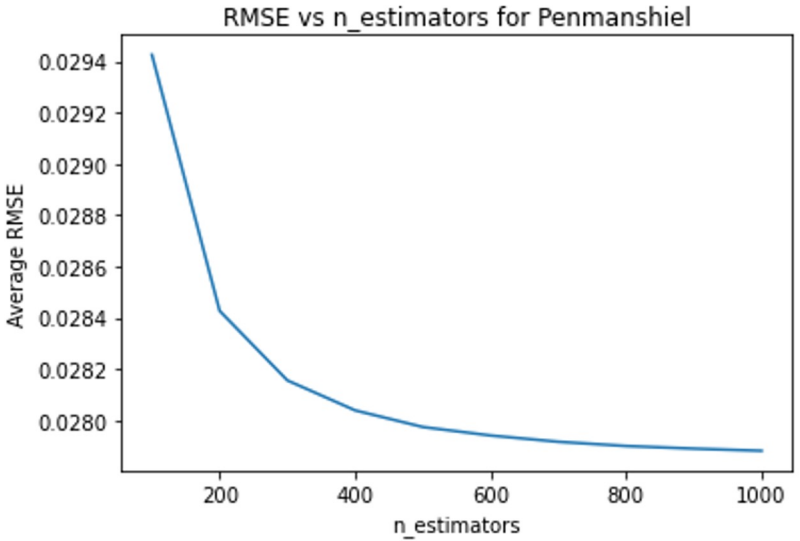
- Further improvement with search grid.
- Further improvement with training the model for each turbine.
- However, run time is going to be an issue.

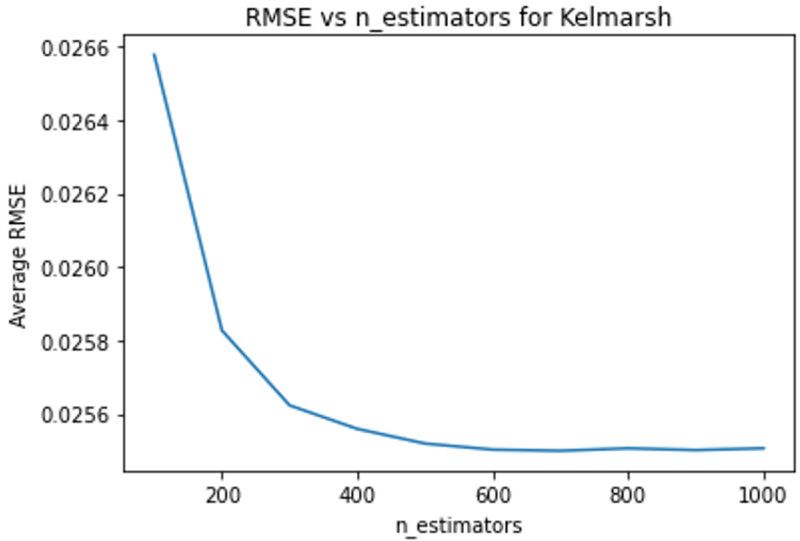
XGBoost: Model

- Same model structure as RandomForest
- One key difference is ensemble method (boosting vs bagging)
- Much faster
- Parameters:
  - 'learning_rate': 0.05, shrinkage of the weights of the new trees that are added to the ensemble during the boosting process. Small values prevent overfitting.
  - 'max_depth': 8, maximum number of levels or nodes that a decision tree can have.
  - 'subsample': 0.7, fraction of observations to be randomly sampled for each tree.
  - 'colsample_bytree': 0.9, controls the fraction of features (columns) to be randomly sampled for each tree in the ensemble
  - 'n_estimators': 400, number of trees

XGBoost: Training & Tuning

- Split 80% training and 20% for validation.
- Train on one all turbines per location: n_estimators only





XGBoost: Results & Remarks

Criteria	MAE	RMSE
XGBoost	0.0146	0.0262
RandomForest	0.0160	0.0281

- Better than RandomForest
- Much faster than random forest
- Expected to perform even better when fully tuned

General Procedure

For All methods we used from DSWE:

1. We used +/- 1 encoding for day/night in both training and test datasets.
2. We used 1:12 encoding for the months in both training and test datasets.

DSWE package in R: Results

Sorted Based on MAE

Methods	Grand MAE
TempGP	0.0162
BART	0.0167
SVM	0.0171
AMK	0.0199
KNN	0.0264

Sorted Based on RMSE

Methods	Grand RMSE
SVM	0.0289
BART	0.0295
TempGP	0.0300
AMK	0.0325
KNN	0.0416

Thank You
For Your Attention