Nihal Patil Follow

Jul 17 · 2 min read · ▶ Listen

🔖 Save
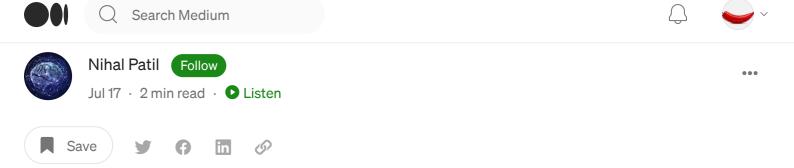
Random Forest cheat sheet

- We have seen in decision tree topic that decision tree tends to overfit a lot that is we use RF.

- Every forest is made up of trees and RF is no exception for that cause RF is made up of DTs.

- Works very well than many other ML algos.

- RF follows Ensemble technique from which it uses Bagging.

- Bagging is nothing but the Bootstrap Aggregation. B from Bootstrap Agg from Aggregation.

- Easy to Parallelize

- OOB error and OOB Score : how our model behaves against testing data.

- Typically 2/3rd of data is row sampled.

- Typically root(features) or log2(features) sampled in feature sampling.

- Hard Voting and Soft Voting. By default SK Learn uses hard voting.

- For Ensemble 2 conditions should be satisfied to accept the Ensembled model:

Diversity : Model should be diverse.

Acceptability : Model should be acceptable enough.

- For both regression and classification.

- For regression it uses median or mean.SK Learn by default uses Mean.

- In DT the model suppresses one of the attribute but, by feature sampling every feature gets equal importance.

- In hyperparameter tuning it comprises of n_estimators' along with all DT hyperparameters.
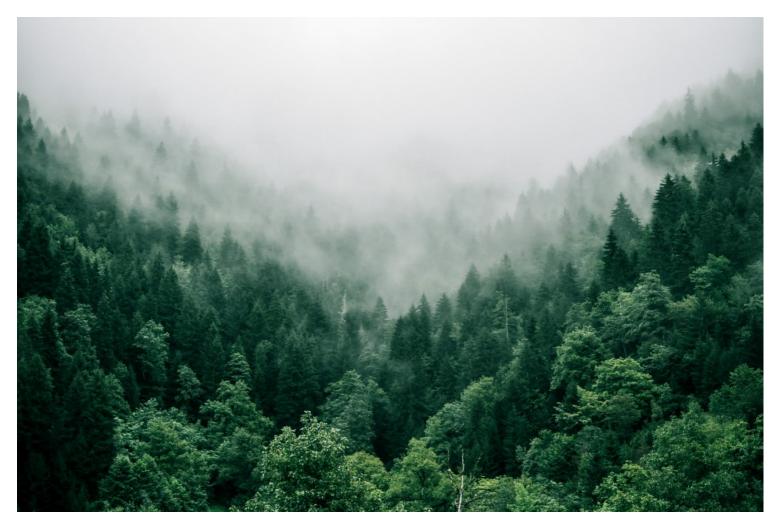
![Foggy forest on a mountainside]

Photo by Marita Kavelashvili on Unsplash

Advantages:

1. It reduces Overfitting of DTs.

2. Doesn't affect by outliers.

3. Non-Parametric.

4. Feature Scaling is not required.

5. It improves the testing accuracy.

6. Regression and classification both.

7. Doesn't suppress the attribute like DTs.

8. Easy to parallelize

9. Stable.

10. Works well on high dimensional data.

Disadvantages:

1. More Computation req.

2. More Time Req.

3. Black Box Model

4. You can't explain its mathematical intuitions in layman's language.

5. Highly Complex

Towards Data Science   Machine Learning   Artificial Intelligence   Random Forest

Data Science