Nihal Patil   Follow

Jul 17 · 2 min read · ▶ Listen

🔖 Save   🐦   📘   in   🔗

Decision Tree cheat sheet

- The Decision Tree is the as like the way we make decisions day-to day life.

- It is one the of supervised ML algo.

- It looks like an if-else ladder code in python.

- tree-based algorithm.

- classification and regression algorithm

- e.g. Doctor treating cancer patient deciding which medicines to give on certain decisions

- Works better for classification.

- Non Parametric

- We prefer over logistic regression when there is a lot of categorial-mixed data.

- No impact of outliers.

- We try to achieve homogeneity in DT.

- To achieve homogeneity we use :

1 Entropy = -[$\sum$ Pi log Pi]

2 Gini = 1- [$\sum$ Pi²]

- By using entropy and Gini we calculate Information Gain(IG) of Decision Node.

- IG = H(s) — wt. avg * (entropy of each feature)

- If IG is maximum then leaf node is tending to be pure.

- Leaf node carries the classifica. ⟨ut.

👏 53 | 💬 | •••

- Decision node used to make decision.

- If IG or Gini has minimum value i. e less impurity then Root or Decision node is best node.

- Decision Tree tends to overfit a lot.

- To overcome overfitting we use regularization

Tree Truncation *: Hyperparameter Tuning : Min sample split, Min sample leaf, max depth*

Tree Pruning : Cost Complexity Pruning i.e. CCP alpha. As CCP alpha values get increases the leaf node will get pruning.
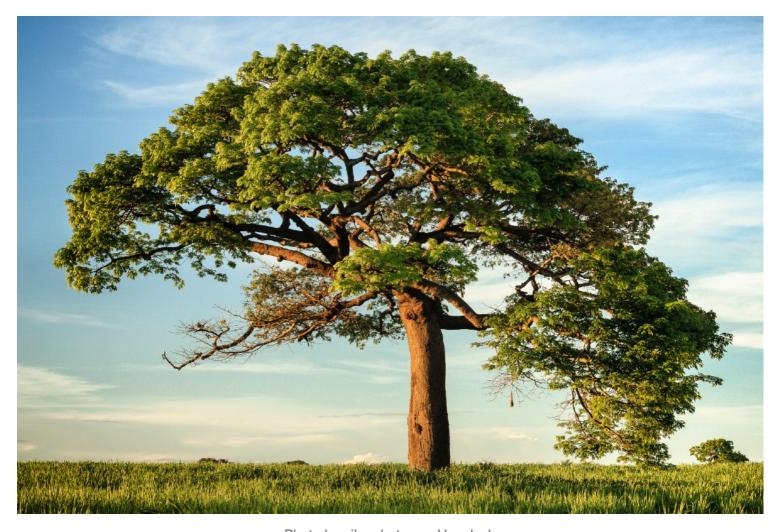

Photo by niko photos on Unsplash

Advantages and Disadvantages

Advantages :

1.No effect of outliers.

2.Classification and regression.

3.Non Parametric

4.Easy to implement and understand.

5.We can visualize the tree.

6.Scaling is not required.

Disadvantages :

1.Tends to overfit a lot.

2. Low Bias and High Variance mostly.

3.Unstable a lot.

ID3 for multiclass.
CART for binary.