

Summary/Review

Decision trees split your data using impurity measures. They are a greedy algorithm and are not based on statistical assumptions.

The most common splitting impurity measures are Entropy and Gini index. Decision trees tend to overfit and to be very sensitive to different data.

Cross validation and pruning sometimes help with some of this.

Great advantages of decision trees are that they are really easy to interpret and require no data preprocessing.