Classification and regression trees (CART) – Decision tree

overfitting?

Variance?

Pruning? – the splitting process results in fully grown trees until the stopping criteria is reached. But, the fully grown tree is likely to overfit data, leading to poor accuracy on unseen data.

Greedy algorithm – always makes the choice that seems to be the best at the moment. This means that it makes a locally-optimal choice in the hope this choice will lead to a globally-optimal solution.

Common decision tree algorithms:

* Giri index
* Ch-square
* Information gain
* Reduction in variance

UseCases: Random Forest, boosting

Java streams and lambdas – 8

Leetcode – w github

Information entropy

entropy