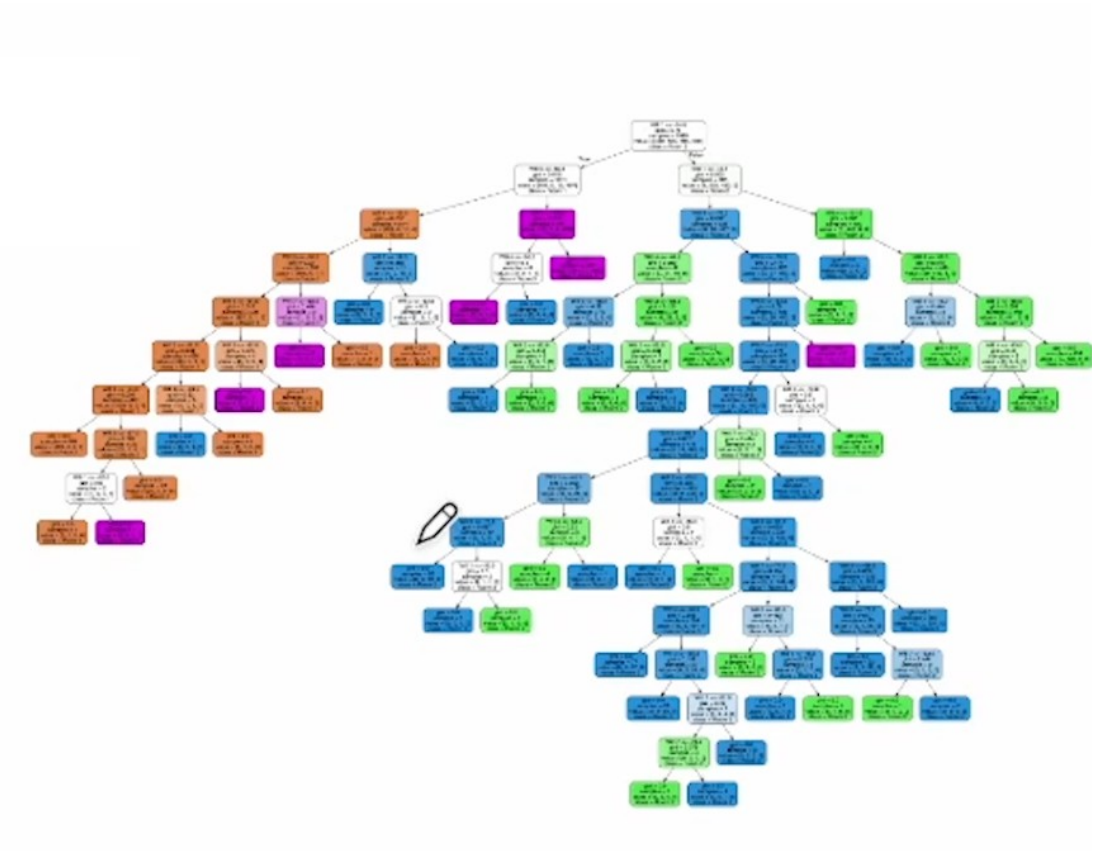
Trees are Unstable

- Trees are very flexible
- A “fully grown” tree is one where all leaves are “pure,” i.e. each leaf contains all + labeled examples and all - labeled examples.
- A fully grown tree has training error zero.
- If the tree is large and the data is limited, the test error of the tree is likely to be high = the tree overfits the data.
- Statisticians say that trees are “high variance” or “unstable”

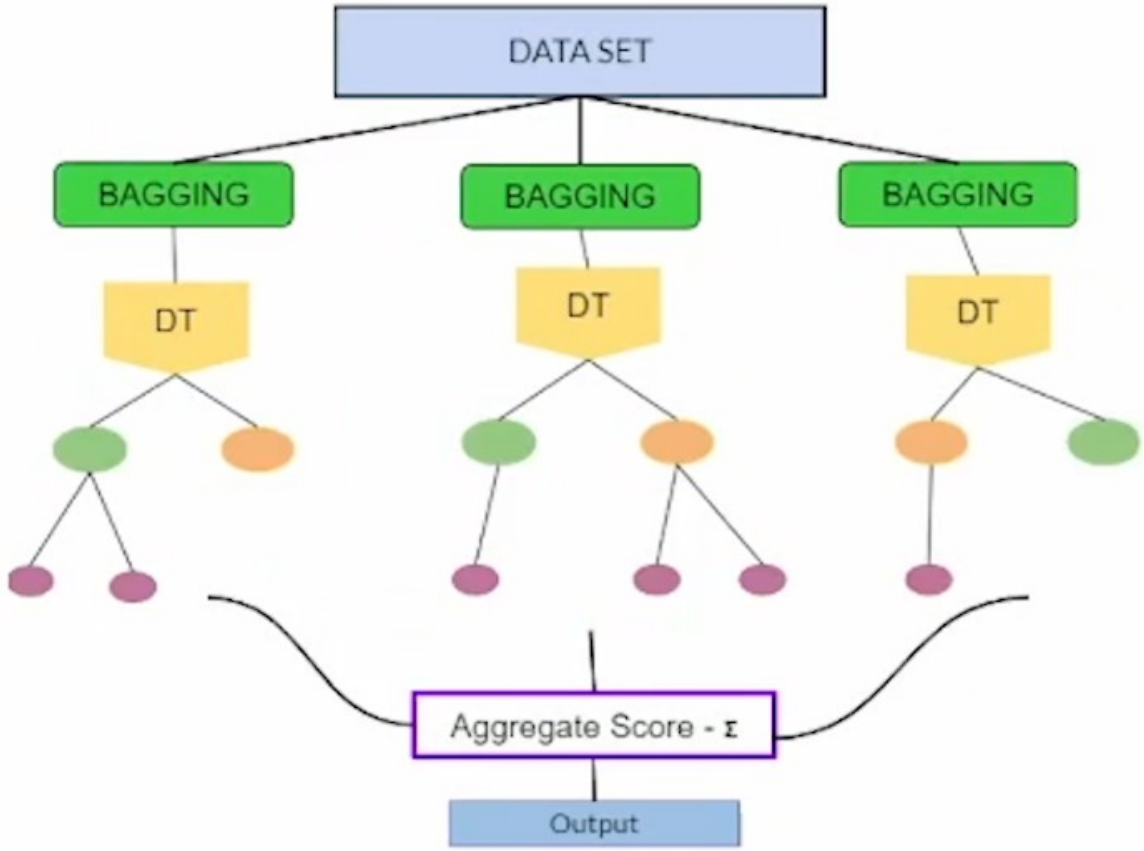
Pruning Trees

- One way to make trees more stable is to prune
- Start by building a tree down to training error of zer
- Iteratively remove leaves until the training error is ϵ =



Bagging Trees

- Bagging, invented by Leo Breiman in the 90s, is a different way to reduce the variance of trees
- Instead of pruning the tree, we generate many trees using randomly selected subsets of the training data.
- We predict using the majority vote over the trees.
- Related to RANDOM FORESTS and to Boosted Trees/XGBoost. Will describe next.

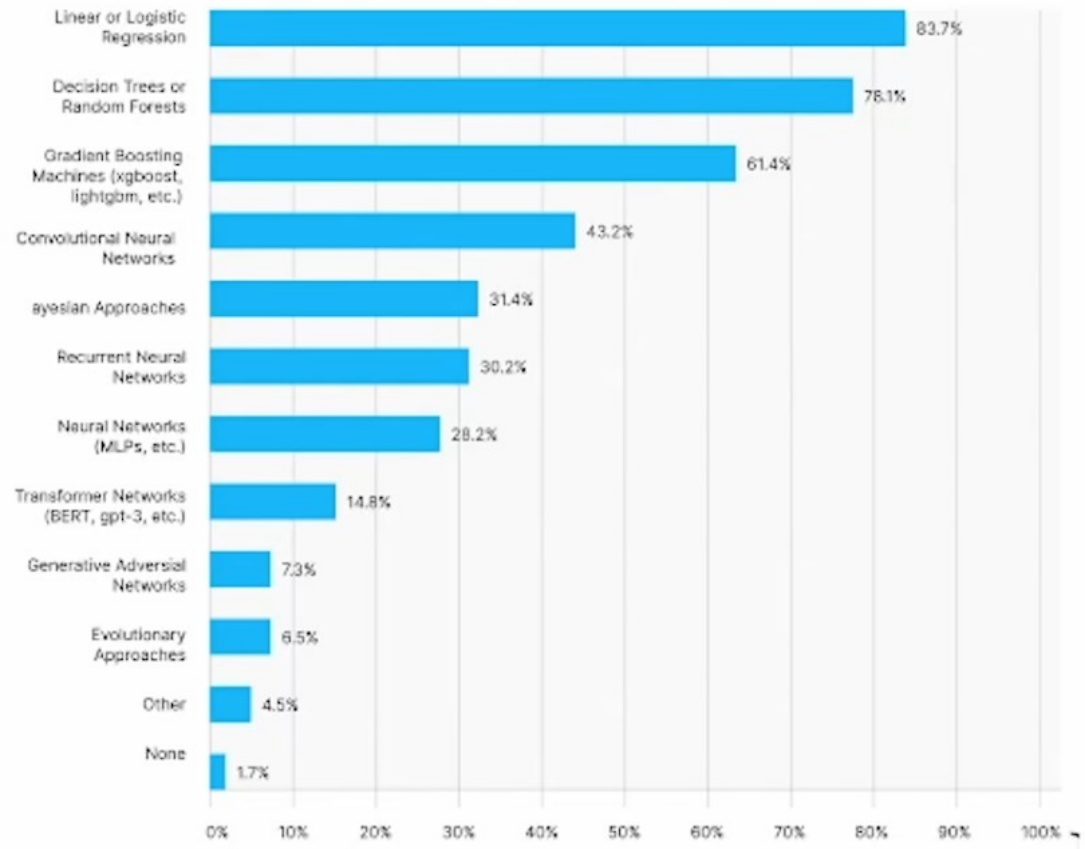


Summary

- Decision trees are a simple and intuitive representation.
- Fully grown trees over-fit the data
- Two ways to reduce over-fitting:
- Pruning trees to make the tree smaller.
- Bagging, Boosting: take the majority vote over many trees.

Are methods based on decision tree still relevant at the age of Deep Neural Networks?

Kaggle Survey of most popular learning algorithms (2020)



Summary

- Decision trees are a simple and intuitive representation.
- Fully grown trees over-fit the data
- Two ways to reduce over-fitting:
- Pruning trees to make the tree smaller.
- Bagging, Boosting: take the majority vote over many trees.