Introduction to Machine Learning

Optimisation is often beneficial to improve the performance of Decision Trees. A deep tree with many leaves will result in overfitting, as each prediction is arising from historical data from a few entities within its leaf. A shallow tree with few leaves will perform poorly because it does not capture many distinctions in the raw data.

Modern decision tree models face this competition between overfitting and underfitting. To combat this, the Random Forest technique can be used. These use many trees, and average the predictions produced by many trees to generate a result.