Random forests are accurate, do not require feature scaling, categorical feature encoding, and little parameter tuning.

Forest is an ensemble of decision trees, usually trained with the bagging method.

A random forest consists of multiple random decision trees. Each tree is built on a random subset of the original data. At each tree node, a subset of features are randomly selected to generate the best split.

A feature’s importance score measures the contribution from the feature. However, the importance of feature doesn’t tell how a feature and the class are correlated.

It is very easy to measure the relative importance of each feature.

In a decision tree each internal node represents a ‘test’ on an attribute. Each branch represents the outcome of the test, and each leaf node represents a class label.

By looking at the feature importance you can decide which featues to drop because they don’t contribute enough.

The more features you have, the more likely your model will suffer from overfitting, and vice versa.

N\_estimators is the number of trees the algorithm build before taking the maximum or average of predictions. In general, a higher number of trees increases performance and makes predictions more stable, but it also slows down computation.

Max\_features: the maximum number of features RF considers to split a node.

Min\_sample\_leaf. A smaller leaf makes the model more prone to capturing noise in train data.