A decision tree combines some decisions, whereas a random forest combines several decision trees. Thus, it is a long process, yet slow.

Whereas, a decision tree is fast and operates easily on large data sets, especially the linear one. The random forest model needs rigorous training. When you are trying to put up a project, you might need more than one model. Thus, a large number of random forests, more the time.

It depends on your requirements. If you have less time to work on a model, you are bound to choose a decision tree. However, stability and reliable predictions are in the basket of random forests.

If you have the passion and want to learn more about artificial intelligence, you can take up [IIIT-B & upGrad’s PG Diploma in Machine Learning and Deep Learning](https://www.upgrad.com/machine-learning-deep-learning-pgc-iiitb/?utm_source=BLOG&utm_medium=BODY&utm_campaign=DV_ML_PGD_BLOG_BODY_92744) that offers 400+ hours of learning, practical sessions, job assistance, and much more.

**What is the difference between the Decision Tree and Random Forest?**

**1. Decision Tree**

Diagram

Description automatically generated

[Source](https://towardsdatascience.com/decision-trees-and-random-forests-df0c3123f991)

Decision Tree is a [supervised learning algorithm](https://www.upgrad.com/blog/types-of-supervised-learning/) used in machine learning. It operated in both classification and regression algorithms. As the name suggests, it is like a tree with nodes. The branches depend on the number of criteria. It splits data into branches like these till it achieves a threshold unit. A decision tree has root nodes, children nodes, and leaf nodes.

Recursion is used for traversing through the nodes. You need no other algorithm. It handles data accurately and works best for a linear pattern. It handles large data easily and takes less time.

**How does it work?**

**1. Splitting**

Data, when provided to the decision tree, undergoes splitting into various categories under branches.

**Must Read: [Naive Bayes Classifier: Pros & Cons, Applications & Types Explained](https://www.upgrad.com/blog/naive-bayes-classifier/)**

**2. Pruning**

Pruning is shredding of those branches furthermore. It works as a classification to subsidize the data in a better way. Like, the same way we say pruning of excess parts, it works the same. The leaf node is reached, and pruning ends. It’s a very important part of decision trees.

**3. Selection of trees**

Now, you have to choose the best tree that can work with your data smoothly.

Here are the factors that need to be considered:

**4. Entropy**

To check the homogeneity of trees, entropy needs to be inferred. If the entropy is zero, it’s homogenous; else not.

**5. Knowledge gain**

Once the entropy is decreased, the information is gained. This information helps to split the branches further.

* You need to calculate the entropy.
* Split the data on the basis of different criteria
* Choose the best information.

Tree depth is an important aspect. The depth informs us of the number of decisions one needs to make before we come up with a conclusion. Shallow depth trees perform better with decision tree algorithms.

**Advantages and Disadvantages of Decision Tree**

**Advantages**

1. Easy
2. Transparent process
3. Handle both numerical and categorical data
4. Larger the data, the better the result
5. Speed

**Disadvantages**

1. May overfit
2. Pruning process large
3. Optimization unguaranteed
4. Complex calculations
5. Deflection high

**Checkout:**[Machine Learning Models Explained](https://www.upgrad.com/blog/machine-learning-models-explained/)

**2. Random Forest**

Graphical user interface, text, application, email

Description automatically generated

[**Source**](https://gist.github.com/dakshtrehan/4bc78644e2a376c1360d8371add41682)

It is also used for supervised learning but is very powerful. It is very widely used.

The basic difference being it does not rely on a singular decision. It assembles randomized decisions based on several decisions and makes the final decision based on the majority.

It does not search for the best prediction. Instead, it makes multiple random predictions. Thus, more diversity is attached, and prediction becomes much smoother.

You can infer Random forest to be a collection of multiple decision trees!

Bagging is the process of establishing random forests while decisions work parallelly.

**1. Bagging**

* Take some training data set
* Make a decision tree
* Repeat the process for a definite period
* Now take the major vote. The one that wins is your decision to take.

**2. Bootstrapping**

Bootstrapping is randomly choosing samples from training data. This is a random procedure.

STEP by STEP

* Random choose conditions
* Calculate the root node
* Split
* Repeat
* You get a forest

**Read : [Naive Bayes Explained](https://www.upgrad.com/blog/naive-bayes-explained/)**

**Advantages and Disadvantages of Random Forest**

**Advantages**

1. Powerful and highly accurate
2. No need to normalizing
3. Can handle several features at once
4. Run trees in parallel ways

**Disadvantages**

1. They are biased to certain features sometimes
2. Slow
3. Can not be used for linear methods
4. Worse for high dimensional data