

## Why Random forest algorithm

To address why random forest algorithm. I am giving you the below advantages.

The same random forest algorithm or the random forest classifier can use for both classification and the regression task. Random forest classifier will handle the missing values. When we have more trees in the forest, random forest classifier won't overfit the model. Can model the random forest classifier for categorical values also. Will discuss these advantages in the random forest algorithm advantages section of this article. Until think through the above advantages of random forest algorithm compared to the other classification algorithms.

Random forest algorithm real life example Random Forest Example Random Forest Example

Before you drive into the technical details about the random forest algorithm. Let's look into a real life example to understand the layman type of random forest algorithm.

Suppose Mady somehow got 2 weeks leave from his office. He wants to spend his 2 weeks by traveling to the different place. He also wants to go to the place he may like.

So he decided to ask his best friend about the places he may like. Then his friend started asking about his past trips. It's just like his best friend will ask, You have been visited the X place did you like it?

Based on the answers which are given by Mady, his best start recommending the place Mady may like. Here his best formed the decision tree with the answer given by Mady.

As his best friend may recommend his best place to Mady as a friend. The model will be biased with the closeness of their friendship. So he decided to ask few more friends to recommend the best place he may like.

Now his friends asked some random questions and each one recommended one place to Mady. Now Mady considered the place which is high votes from his friends as the final place to visit.

In the above Mady trip planning, two main interesting algorithms decision tree algorithm and random forest algorithm used. I hope you find it already. Anyhow, I would like to highlight it again.

## Decision Tree:

- To recommend the best place to Mady, his best friend asked some questions. Based on the answers given by mady, he recommended a place. This is decision tree algorithm approach. Will explain why it is a decision tree algorithm approach.
- Mady friend used the answers given by mady to create rules. Later he used the created rules to recommend the best place which mady will like. These rules could be, mady like a place with lots of tree or waterfalls ..etc
- In the above approach mady best friend is the decision tree. The vote (recommended place) is the leaf of the decision tree (Target class). The target is finalized by a single person, In a technical way of saying, using an only single decision tree.

## Random Forest Algorithm:

In the other case when mady asked his friends to recommend the best place to visit. Each friend asked him different questions and come up their recommend a place to visit. Later

mady consider all the recommendations and calculated the votes. Votes basically is to pick the popular place from the recommend places from all his friends.

Mady will consider each recommended place and if the same place recommended by some other place he will increase the count. At the end the high count place where mady will go.

In this case, the recommended place (Target Prediction) is considered by many friends. Each friend is the tree and the combined all friends will form the forest. This forest is the random forest. As each friend asked random questions to recommend the best place visit.

Now lets use the above example to understand how the random forest algorithm work.

## How Random forest algorithm works

How random forest algorithm works How random forest algorithm works

Lets look at the pseudocode for random forest algorithm and later we can walk through each step in the random forest algorithm.

The pseudocode for random forest algorithm can split into two stages.

Random forest creation pseudocode. Pseudocode to perform prediction from the created random forest classifier. First, lets begin with random forest creation pseudocode

Random Forest pseudocode: Randomly select  $k$  features from total  $m$  features.

Where  $k \ll m$  Among the  $k$  features, calculate the node  $d$  using the best split point.

Split the node into daughter nodes using the best split.

Repeat 1 to 3 steps until  $l$  number of nodes has been reached.

Build forest by repeating steps 1 to 4 for  $n$  number times to create  $n$  number of trees.

The beginning of random forest algorithm starts with randomly selecting  $k$  features out of total  $m$  features. In the image, you can observe that we are randomly taking features and observations.

In the next stage, we are using the randomly selected  $k$  features to find the root node by using the best split approach.

The next stage, We will be calculating the daughter nodes using the same best split approach. Will the first 3 stages until we form the tree with a root node and having the target as the leaf node.

Finally, we repeat 1 to 4 stages to create  $n$  randomly created trees. This randomly created trees forms the random forest.