

Random Forest using R

Spoken Tutorial Project

<https://spoken-tutorial.org>

National Mission on Education through ICT

<https://sakshat.ac.in>

Tanmay Srinath

Madhuri Ganapathi

IIT Bombay

8 June 2022



Learning Objectives



Learning Objectives

We will learn about:



Learning Objectives

We will learn about:

▶ **Random Forest**



Learning Objectives

We will learn about:

- ▶ Random Forest
- ▶ Bagging



Learning Objectives

We will learn about:

- ▶ **Random Forest**
- ▶ **Bagging**
- ▶ **Benefits of Random Forest**



Learning Objectives

We will learn about:

- ▶ **Random Forest**
- ▶ **Bagging**
- ▶ **Benefits of Random Forest**
- ▶ **Applications of Random Forest**



Learning Objectives

We will learn about:

- ▶ **Random Forest**
- ▶ **Bagging**
- ▶ **Benefits of Random Forest**
- ▶ **Applications of Random Forest**
- ▶ **Random Forest on iris dataset**



Learning Objectives

We will learn about:

- ▶ **Random Forest**
- ▶ **Bagging**
- ▶ **Benefits of Random Forest**
- ▶ **Applications of Random Forest**
- ▶ **Random Forest on iris dataset**
- ▶ **Tuning a Random Forest model**



System Specifications



System Specifications

- ▶ **Ubuntu Linux OS version 20.04**



System Specifications

- ▶ **Ubuntu Linux OS version 20.04**
- ▶ **R version 4.2.0**



System Specifications

- ▶ **Ubuntu Linux OS version 20.04**
- ▶ **R version 4.2.0**
- ▶ **RStudio version 2022.02.3**



Pre-requisites



Pre-requisites

▶ Basics of R Programming



Pre-requisites

- ▶ **Basics of R Programming**
- ▶ **Basics of Machine Learning**



Pre-requisites

- ▶ **Basics of R Programming**
- ▶ **Basics of Machine Learning**



Pre-requisites

- ▶ Basics of R Programming
- ▶ Basics of Machine Learning

If not, please access the relevant tutorials on

<https://spoken-tutorial.org/>



Random Forest



Random Forest

- ▶ It is a powerful and versatile supervised machine learning algorithm



Random Forest

- ▶ It grows and combines multiple decision trees to create a "forest"



Random Forest

- ▶ It grows and combines multiple decision trees to create a "forest"
- ▶ It can be used for both classification and regression problems



Benefits of Random Forest



Benefits of Random Forest

- ▶ **Random forests are created from subsets of data**



Benefits of Random Forest

- ▶ Random forests are created from subsets of data
- ▶ The final output is based on average or majority ranking



Benefits of Random Forest

- ▶ Random forests are created from subsets of data
- ▶ The final output is based on average or majority ranking
- ▶ Hence the problem of overfitting is taken care of



Bagging

- ▶ **Random Forest uses a concept called bagging to improve its performance**



Bagging



Bagging

- ▶ It is used to reduce the variance of statistical learning methods



Bagging

- ▶ It is used to reduce the variance of statistical learning methods
- ▶ It works by creating multiple decision trees on multiple bootstrapped datasets



Bagging

- ▶ Each of these decision trees are deep and not pruned



Bagging

- ▶ Each of these decision trees are deep and not pruned
- ▶ The results from these trees are averaged to provide the final output



Bagging in Random Forests



Bagging in Random Forests

- ▶ Bagging process is used to decorrelate the trees that make up a random forest



Bagging in Random Forests

- ▶ A random sample of predictors are chosen for each split in the decision tree



Bagging in Random Forests

- ▶ A random sample of predictors are chosen for each split in the decision tree
- ▶ Thus, the average of the trees will be less variable and more reliable



Applications of Random Forest



Applications of Random Forest

- ▶ It is used in customer segmentation

<https://archive.ics.uci.edu/ml/datasets/Online+Retail+II>



Applications of Random Forest

- ▶ It is used in cancer diagnosis
[https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+\(diagnostic\)](https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(diagnostic))



Random Forest

Let us implement Random Forest on the iris dataset



Download Files



Download Files

We will use:



Download Files

We will use:

- ▶ A script file `RandomForest.R`



Download Files

We will use:

- ▶ A script file **RandomForest.R**



Download Files

We will use:

- ▶ A script file **RandomForest.R**

Download this file from the **Code files** link of this tutorial

Make a copy and then use it for practising



Tuning a Random Forest



Tuning a Random Forest

- ▶ Sometimes the default parameters of the model are not optimal



Tuning a Random Forest

- ▶ Sometimes the default parameters of the model are not optimal
- ▶ Thus we need to tune our Random Forest by changing a few parameters



Tuning a Random Forest

- ▶ Sometimes the default parameters of the model are not optimal
- ▶ Thus we need to tune our Random Forest by changing a few parameters
- ▶ In R, this is done using the `tuneRF()` function



Summary

We have learnt about:

- ▶ **Random Forest**
- ▶ **Bagging**
- ▶ **Advantages of Random Forest**
- ▶ **Applications of Random Forest**
- ▶ **Random Forest on iris dataset**
- ▶ **Tuning a Random Forest model**



Assignment



Assignment

- ▶ **Create a Random Forest for PimaIndiansDiabetes dataset**



Assignment

- ▶ **Create a Random Forest for PimaIndiansDiabetes dataset**
- ▶ **Tune the model using tuneRF() command**



About the Spoken Tutorial Project

- ▶ Watch the video available at https://spoken-tutorial.org/What_is_a_Spoken_Tutorial
- ▶ It summarises the Spoken Tutorial project
- ▶ If you do not have good bandwidth, you can download and watch it



Spoken Tutorial Workshops

The Spoken Tutorial Project Team

- ▶ Conducts workshops using spoken tutorials
- ▶ Gives certificates to those who pass an online test
- ▶ For more details, please write to contact@spoken-tutorial.org



Answers for THIS Spoken Tutorial

- ▶ **Questions in THIS Spoken Tutorial?**
- ▶ **Visit**
<https://forums.spoken-tutorial.org>
- ▶ **Choose the minute and second where you have the question**
- ▶ **Explain your question briefly**
- ▶ **The FOSSEE project will ensure an answer**

You will have to register to ask questions



Forum to answer questions

- ▶ Questions not related to the Spoken Tutorial?
- ▶ Do you have general/technical questions on the Software?
- ▶ Please visit the FOSSEE Forum
<https://forums.fossee.in/>
- ▶ Choose the Software and post your question



Textbook Companion Project

- ▶ The FOSSEE team coordinates the coding of solved examples of popular books and case study projects
- ▶ We give certificates to those who do this

For more details, please visit these sites:

<https://r.fossee.in/>
<https://fossee.in/>



Acknowledgements

- ▶ **The Spoken Tutorial and FOSSEE projects are funded by the Ministry of Education, Govt. of India**



About the Contributors

- ▶ **This tutorial is contributed by Tanmay Srinath and Madhuri Ganapathi, IIT Bombay**

