

Decision Tree 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

6 June 2022



Learning Objectives



Learning Objectives

We will learn about:



Learning Objectives

We will learn about:

▶ **Decision Tree**



Learning Objectives

We will learn about:

- ▶ Decision Tree
- ▶ Advantages of Decision Trees



Learning Objectives

We will learn about:

- ▶ Decision Tree
- ▶ Advantages of Decision Trees
- ▶ Applications of Decision Trees



Learning Objectives

We will learn about:

- ▶ Decision Tree
- ▶ Advantages of Decision Trees
- ▶ Applications of Decision Trees
- ▶ Practical implementation of Decision Trees in R



Learning Objectives

We will learn about:

- ▶ Decision Tree
- ▶ Advantages of Decision Trees
- ▶ Applications of Decision Trees
- ▶ Practical implementation of Decision Trees in R
- ▶ Disadvantages of Decision Trees



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/>



Decision Tree



Decision Tree

- ▶ It is a visualisation technique



Decision Tree

- ▶ It is a visualisation technique
- ▶ It can be used for both classification and regression tasks



Decision Tree

- ▶ It is a visualisation technique
- ▶ It can be used for both classification and regression tasks
- ▶ It works on non-gaussian or discrete data



Advantages of Decision Tree



Advantages of Decision Tree

- ▶ **Decision trees are easy to understand**



Advantages of Decision Tree

- ▶ Decision trees are easy to understand
- ▶ They are mirrors of human decision making



Applications of Decision Tree



Applications of Decision Tree

- ▶ Predicting salaries of the employees

<https://www.rdocumentation.org/packages/vcd/versions/1.4-8/topics/Hitters>



Applications of Decision Tree

- ▶ **Diagnosis of diseases and ailments**
<https://archive.ics.uci.edu/ml/datasets/Heart+Disease>



Decision Tree



Decision Tree

- ▶ We will construct a **Decision Tree** on the **built-in iris** dataset



Decision Tree

- ▶ We will construct a **Decision Tree** on the **built-in** `iris` dataset
- ▶ It will be used to **predict the species** of a given data



Download Files

We will use:



Download Files

We will use:

- ▶ A script file `DecisionTree.R`



Download Files

We will use:

- ▶ A script file **DecisionTree.R**



Download Files

We will use:

- ▶ A script file **DecisionTree.R**

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

Make a copy and then use it for practising



Disadvantages of Decision Tree



Disadvantages of Decision Tree

- ▶ **Decision trees have reduced predictive power**



Disadvantages of Decision Tree

- ▶ Decision trees have reduced predictive power
- ▶ They can be non-robust



Summary

We have learnt about:

- ▶ **Decision Tree**
- ▶ **Advantages of Decision Trees**
- ▶ **Applications of Decision Trees**
- ▶ **Practical implementation of Decision Trees in R**
- ▶ **Disadvantages of Decision Trees**



Assignment



Assignment

- ▶ **Perform Decision Tree on PimaIndiansDiabetes dataset**



Assignment

- ▶ **Perform Decision Tree on PimaIndiansDiabetes dataset**
- ▶ **Install and import the mlbench package**



Assignment

- ▶ **Perform Decision Tree on**
PimaIndiansDiabetes **dataset**
- ▶ **Install and import the mlbench**
package
- ▶ **Run the data(PimaIndiansDiabetes2)**
command to load the dataset



Assignment

- ▶ **Perform Decision Tree on**
PimaIndiansDiabetes **dataset**
- ▶ **Install and import the mlbench**
package
- ▶ **Run the data(PimaIndiansDiabetes2)**
command to load the dataset
- ▶ **Tabulate the results**



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**

