

# Decision Tree in R

Spoken Tutorial Project

<https://spoken-tutorial.org>

National Mission on Education through ICT

Debatosh Chakraborty

Yate A Ronald

FOSSEE, IIT Bombay

7 June 2024



# Learning Objectives



# Learning Objectives

**We will learn about:**



# Learning Objectives

**We will learn about:**

- **Decision Tree**



# Learning Objectives

We will learn about:

- Decision Tree
- Assumptions for Decision Tree



# Learning Objectives

**We will learn about:**

- **Decision Tree**
- **Assumptions for Decision Tree**
- **Advantages of Decision Tree**



# Learning Objectives

We will learn about:

- Decision Tree
- Assumptions for Decision Tree
- Advantages of Decision Tree
- Implementation of Decision Tree in R



# Learning Objectives

- Plotting the decision tree model



# Learning Objectives

- Plotting the decision tree model
- Evaluation of the model



# Learning Objectives

- Plotting the decision tree model
- Evaluation of the model
- Visualizing the model decision boundary



# Learning Objectives

- Plotting the decision tree model
- Evaluation of the model
- Visualizing the model decision boundary
- Limitations of Decision Tree



# System Specifications



# System Specifications

- Windows 11



# System Specifications

- **Windows 11**
- **R v 4.3.0**



# System Specifications

- Windows 11
- R v 4.3.0
- RStudio v 2023.06.1



# System Specifications

- Windows 11
- R v 4.3.0
- RStudio v 2023.06.1



# System Specifications

- **Windows 11**
- **R v 4.3.0**
- **RStudio v 2023.06.1**

**It is recommended to install R version 4.2.0 or higher**



# Pre-requisites



# Pre-requisites

**To follow this tutorial, learner should know:**



# Pre-requisites

To follow this tutorial, learner should know:

- Basic programming in R



# Pre-requisites

To follow this tutorial, learner should know:

- Basic programming in R
- Basics of Machine Learning



# Pre-requisites

To follow this tutorial, learner should know:

- Basic programming in R
- Basics of Machine Learning
- If not, please access the relevant tutorials on this website  
<https://spoken-tutorial.org>



# What is a Decision Tree?



# What is a Decision Tree?

- It uses a binary tree to split the feature space into several sub-regions



# What is a Decision Tree?

- It uses a binary tree to split the feature space into several sub-regions
- The nodes of the tree are the locations at which the feature space splits



# What is a Decision Tree?

- **Misclassification error, Gini index, and entropy aid in identifying ideal splits**



# What is a Decision Tree?

- Misclassification error, Gini index, and entropy aid in identifying ideal splits
- The decision boundaries in the Decision Tree model are non-linear



# Assumptions for Decision Tree

- The model does not assume any specific distribution of features
- Each observation is independent



# Advantages of Decision Tree



# Advantages of Decision Tree

- It does not require feature variables to be necessarily continuous



# Advantages of Decision Tree

- It does not require feature variables to be necessarily continuous
- Decision trees are intuitive and easy to visualize



# Advantages of Decision Tree

- It does not require feature variables to be necessarily continuous
- Decision trees are intuitive and easy to visualize
- When the response is continuous, the decision tree methodology can be easily implemented as a regression tree



# Implementation of Decision Tree



# Implementation of Decision Tree

- We will construct a Decision Tree on the raisin dataset with two chosen variables



# 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**
- **Raisin Dataset 'raisin.xlsx'**



# Download Files

We will use:

- A script file **DecisionTree.R**
- Raisin Dataset 'raisin.xlsx'



# Download Files

We will use:

- A script file **DecisionTree.R**
- Raisin Dataset 'raisin.xlsx'

Please download these files from the **Code files** link of this tutorial

Make a copy and then use them while practising



# Limitations of Decision Tree

- If the tree is too complex, it can overfit data
- Small variations in data can result in a different tree
- Large trees are difficult to interpret
- Noisy data may cause inaccurate splits



# Summary

**In this tutorial we have learnt about:**

- **Decision Tree**
- **Assumptions for Decision Tree**
- **Advantages of Decision Tree**
- **Implementation of Decision Tree in R**



# Summary

- Plotting the decision tree model
- Evaluation of the model
- Visualizing the model decision boundary
- Limitations of Decision Tree



# Assignment



# Assignment

- Apply Decision Tree on PimaIndiansDiabetes dataset
- Install the **pdp** package and import the dataset using the **data(pima)** command
- Visualize the decision tree and measure the accuracy of the model



# About the Spoken Tutorial Project

- Watch the video available at [https://spoken-tutorial.org/What\\_is\\_a\\_Spoken\\_Tutorial](https://spoken-tutorial.org/What_is_a_Spoken_Tutorial)
- It summarises the Spoken Tutorial project



# About the Spoken Tutorial Project

- Watch the video available at [https://spoken-tutorial.org/What\\_is\\_a\\_Spoken\\_Tutorial](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](mailto: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 Spoken Tutorial project will ensure an answer



# Forum to answer questions

- Do you have any general/technical questions?
- Please visit the forum given in the link  
<https://forums.fossee.in/>



# 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

- **Spoken Tutorial project was established by the Ministry of Education(MoE), Govt of India**



# Thank You

- This tutorial is contributed by **Debatosh Chakraborty and Yate A Ronald O** from **IIT Bombay**
- Thank you for joining

