Decision tree implementation in R

There are various ways to implement the decision tree algorithm. One of the ways to implement a decision tree is using the rpart library.

The dataste is collected from the Kaggle website. It contains information about the patients who are having coronary heart disease.

The link to the dataset is https://www.kaggle.com/code/captainozlem/framingham-chd-preprossing-data/data

Prerequisite to run this code

**You have to install two packages from the cran repository.**

**install.packages(rpart)– This library is used to implement a decision tree**

**install.packages(rpart.plot)– This library is used to plot the decision tree graph.**

**install.packages(caTools)– This library is used to split the dataset into training and testing**

Import the required libraries

**library(rpart)**

**library(rpart.plot)**

**library(caTools)**

**Load the CSV dataset**

**data <- read.csv('CHD\_preprocessed.csv')**

Basic data exploration

**#checking for the shape of the data**

**ndim(data)**

**# checking for the column in the dataset**

**str(data)**

**# summary statistics of the dataset**

**summary(data)**

Splitting the dataset

**set.seed(123)**

**#split the dataset into training and testing**

**split = sample.split(df$TenYearCHD, SplitRatio = 0.75)**

**training\_set = subset(df, split == TRUE)**

**test\_set = subset(df, split == FALSE)'**

Applying decision tree algorithms

**tree <- rpart(TenYearCHD ~ ., data = training\_set,method = 'class')**

**rpart.plot(tree)**