## Feature Selection with Information Gain (IG)

Information Gain measures the reduction in class label entropy achieved by splitting on a feature. Features that yield larger entropy reduction are considered more informative for predicting the target.

This example shows how to use the Information Gain method to rank attributes and select a relevant subset for the target. Then, we apply the transformation to keep only the chosen attributes along with the target.

Prerequisites - R packages: daltoolbox, daltoolboxdp

# Installation (if needed)  
#install.packages("daltoolboxdp")

# Loading packages  
library(daltoolbox)  
library(daltoolboxdp)

# Example data  
iris <- datasets::iris

# Information Gain (IG) - step by step  
  
# 1) Fit the feature selector (target: Species)  
myfeature <- fit(fs\_ig("Species"), iris)  
  
# 2) View selected features  
print(myfeature$features)

## [1] "Petal.Width" "Petal.Length"

# 3) Apply transformation to keep only selected features + target  
data <- transform(myfeature, iris)  
print(head(data))

## Petal.Width Petal.Length Species  
## 1 0.2 1.4 setosa  
## 2 0.2 1.4 setosa  
## 3 0.2 1.3 setosa  
## 4 0.2 1.5 setosa  
## 5 0.2 1.4 setosa  
## 6 0.4 1.7 setosa

References - Quinlan, J. R. (1993). C4.5: Programs for Machine Learning. Morgan Kaufmann.