## Class Balance: Subsampling

Random undersampling reduces the majority class by removing examples until the classes are balanced. It can mitigate bias toward the majority class but may discard useful information if too aggressive.

This example shows how to handle class imbalance by applying subsampling (reduce the majority class) on an imbalanced subset of the Iris dataset.

Prerequisites - R packages: daltoolbox, daltoolboxdp

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

# Loading packages  
library(daltoolbox)  
library(daltoolboxdp)

# Example data and creation of artificial imbalance  
iris <- datasets::iris  
data(iris)  
mod\_iris <- iris[c(1:50,51:71,101:111),]  
table(mod\_iris$Species) # original distribution

##   
## setosa versicolor virginica   
## 50 21 11

# Subsampling - reduce the majority class to balance  
bal <- bal\_subsampling('Species')  
bal <- daltoolbox::fit(bal, mod\_iris)  
adjust\_iris <- daltoolbox::transform(bal, mod\_iris)  
table(adjust\_iris$Species) # distribution after subsampling

##   
## setosa versicolor virginica   
## 11 11 11

References - He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9), 1263–1284.