## Class Balance: Oversampling

Random oversampling increases the representation of the minority class by replicating existing minority samples until a desired class balance is reached. It is simple and effective but may increase overfitting by duplicating examples.

This example shows how to handle class imbalance by applying oversampling (increase the minority 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),] # subset with imbalanced classes  
table(mod\_iris$Species) # original distribution

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

# Oversampling - increase the minority class to balance  
bal <- bal\_oversampling('Species')  
bal <- daltoolbox::fit(bal, mod\_iris)  
adjust\_iris <- daltoolbox::transform(bal, mod\_iris)  
table(adjust\_iris$Species) # distribution after oversampling

##   
## setosa versicolor virginica   
## 50 42 44

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