

Jakub Kosterna  
04.03.2021

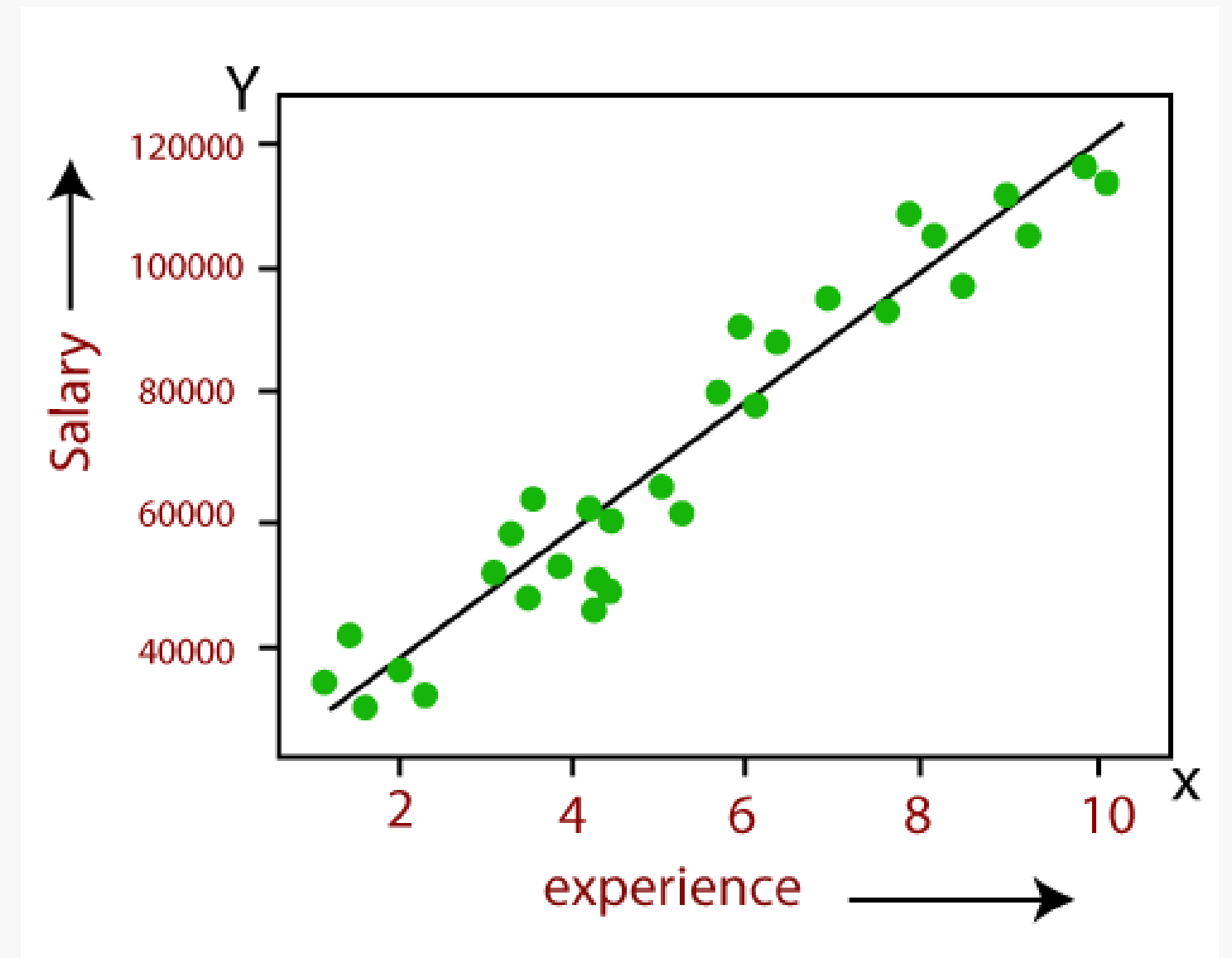
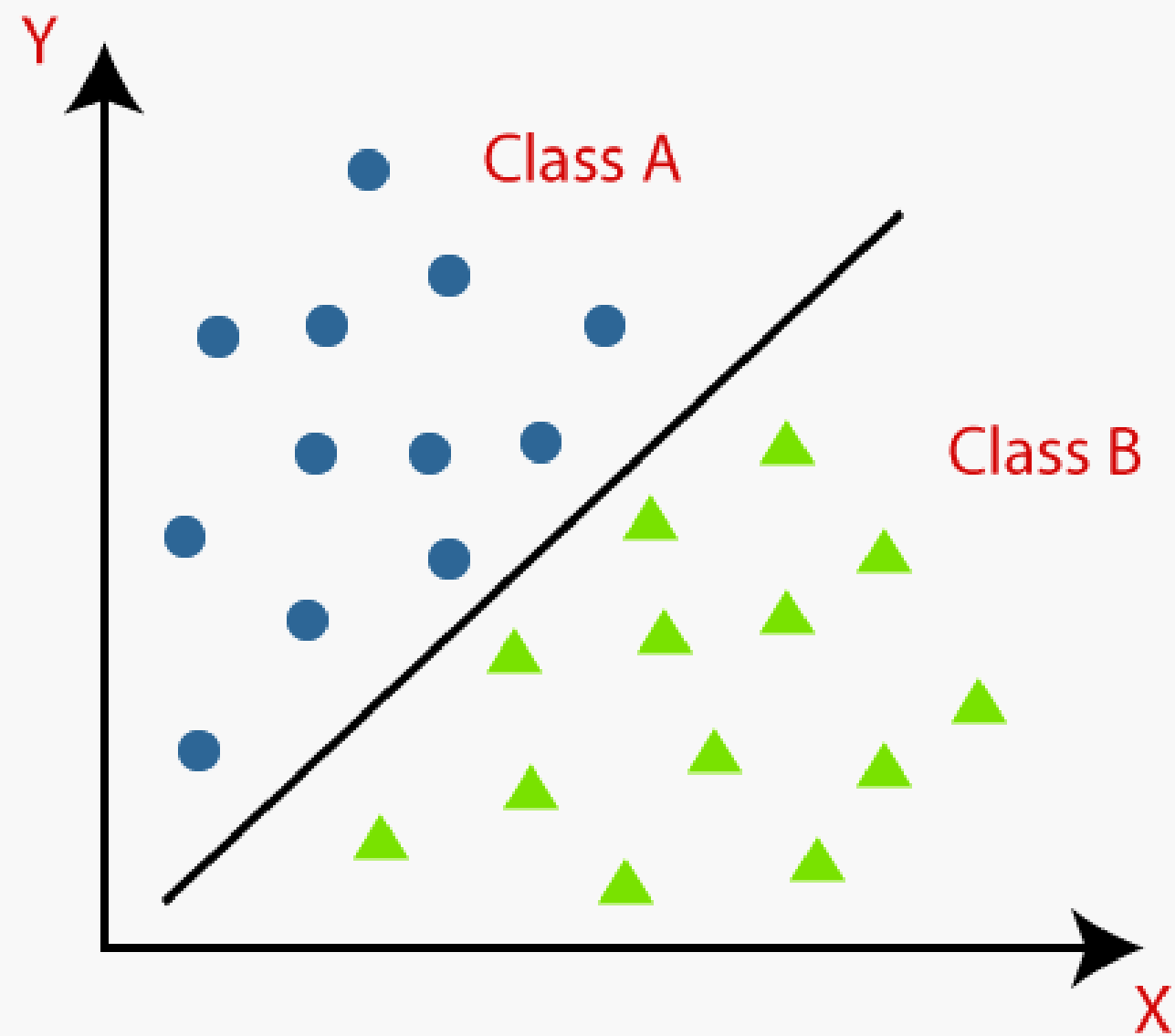
# XGBOOST

EXTREME GRADIENT BOOSTING

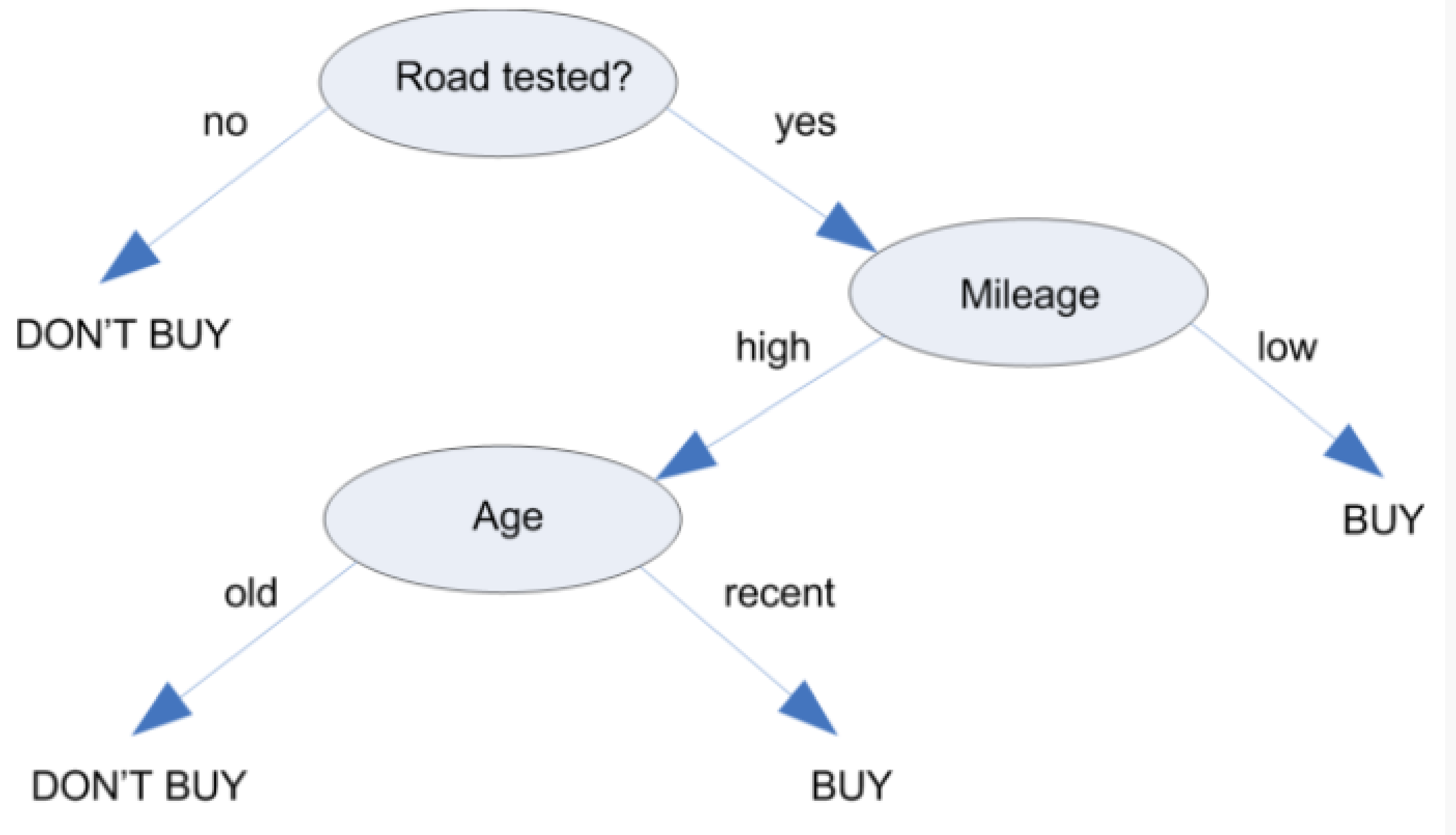
- 2014
- C++
- Higgs Machine Learning Challenge
- Python: *scikit-learn*
- R: *caret*
- + Java, Scala, Julia...

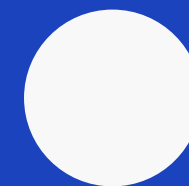
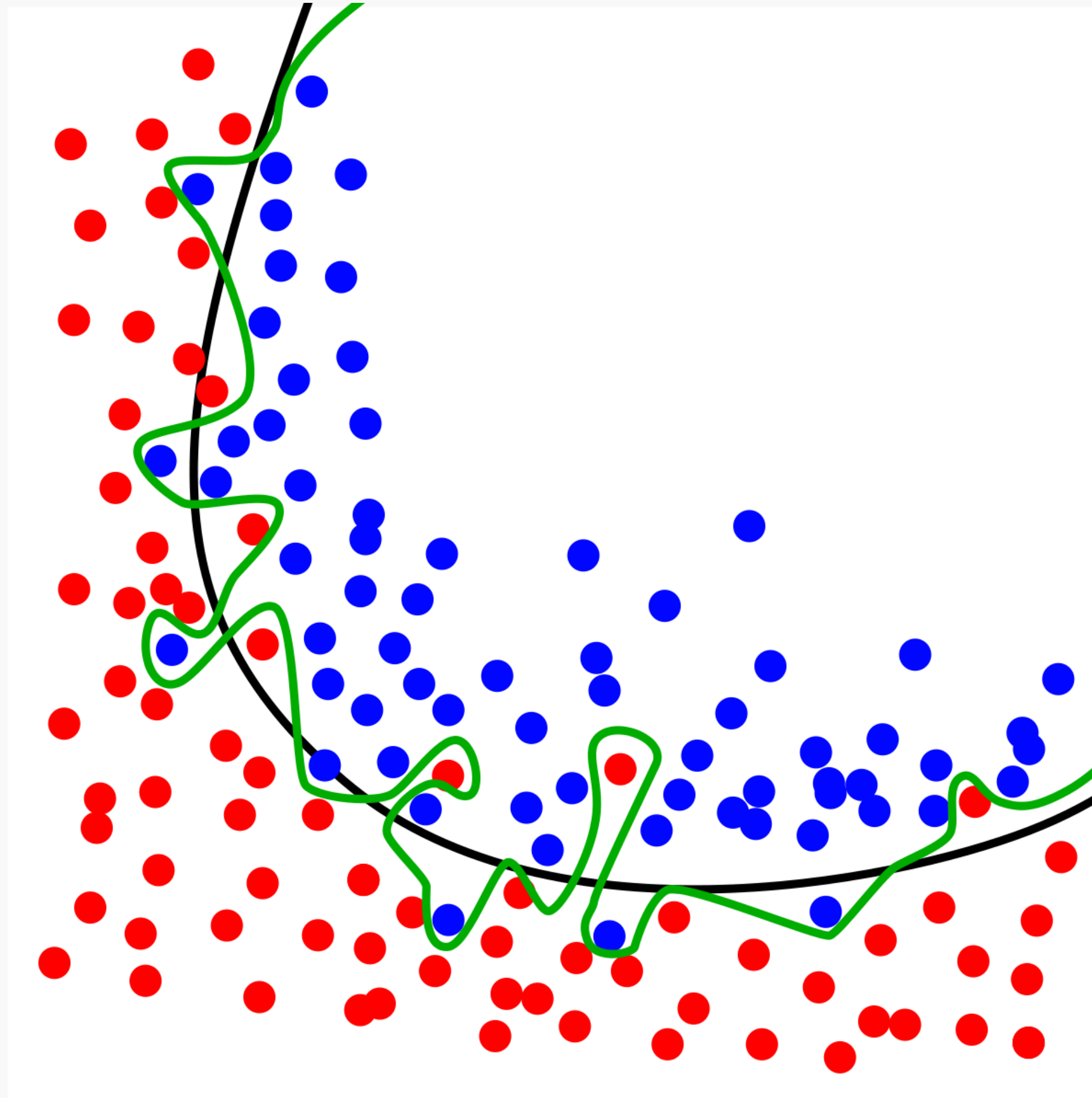


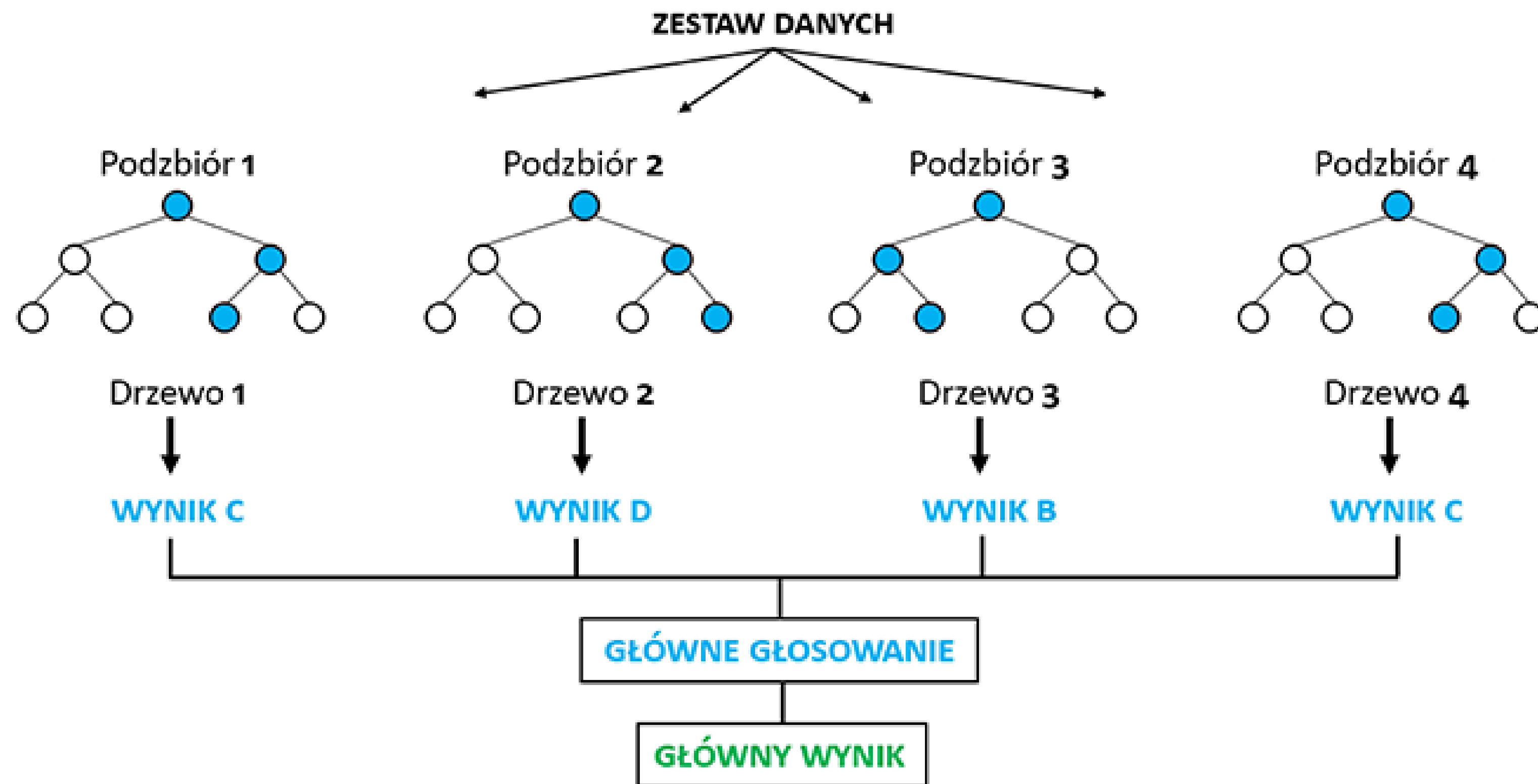
# Klasyfikacja czy regresja?





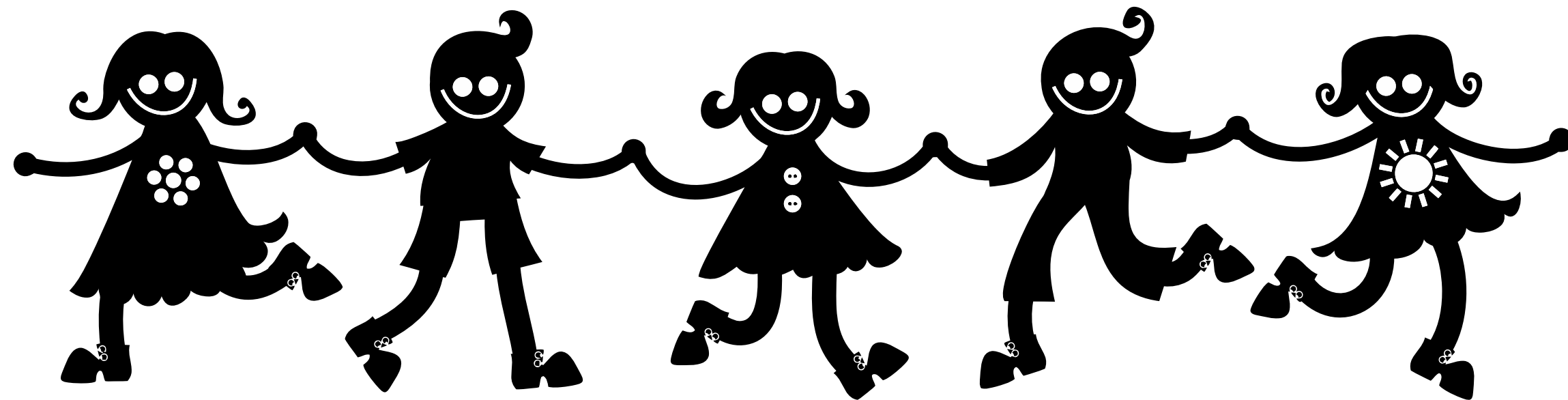








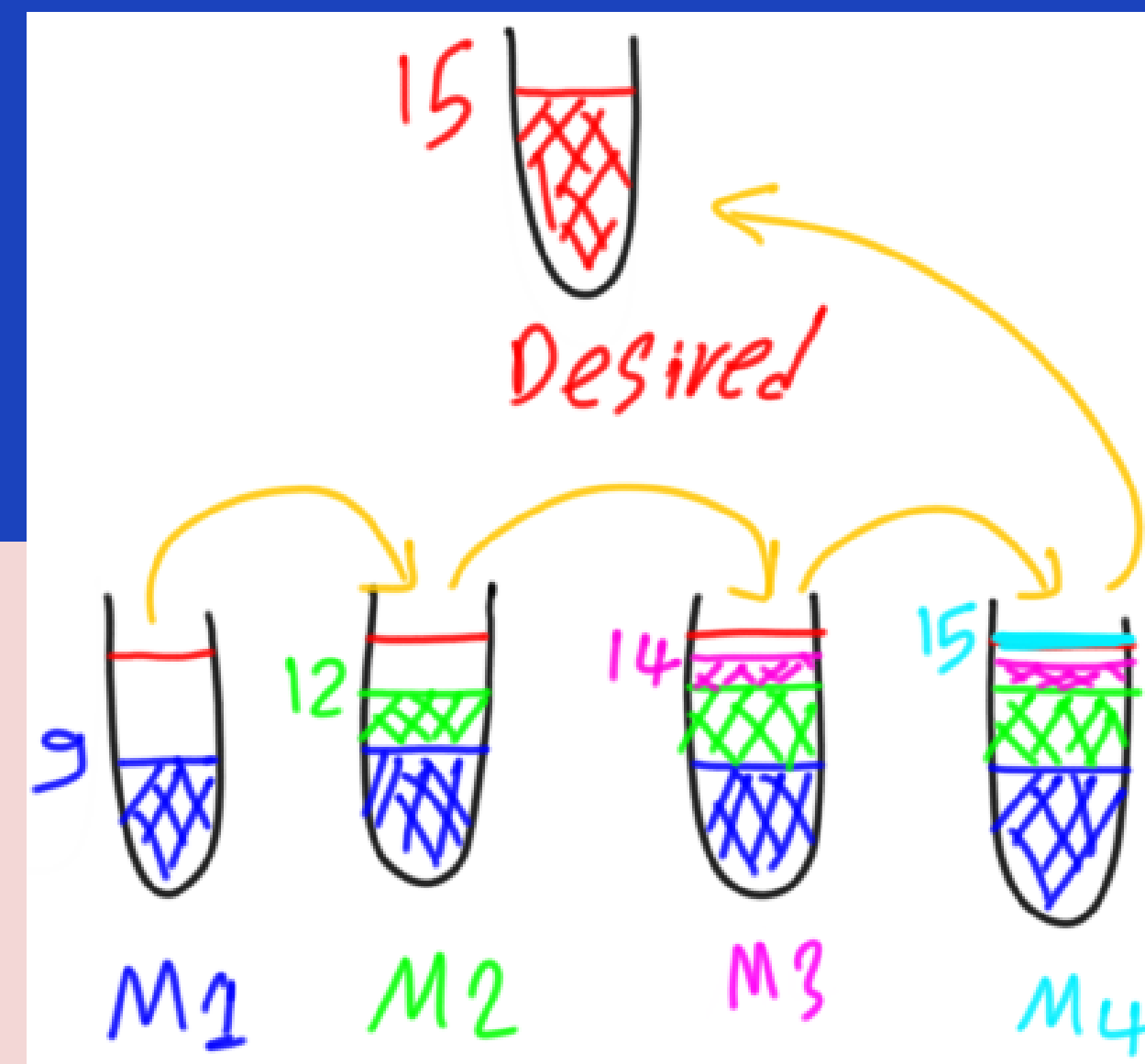
# AdaBoost (2003)

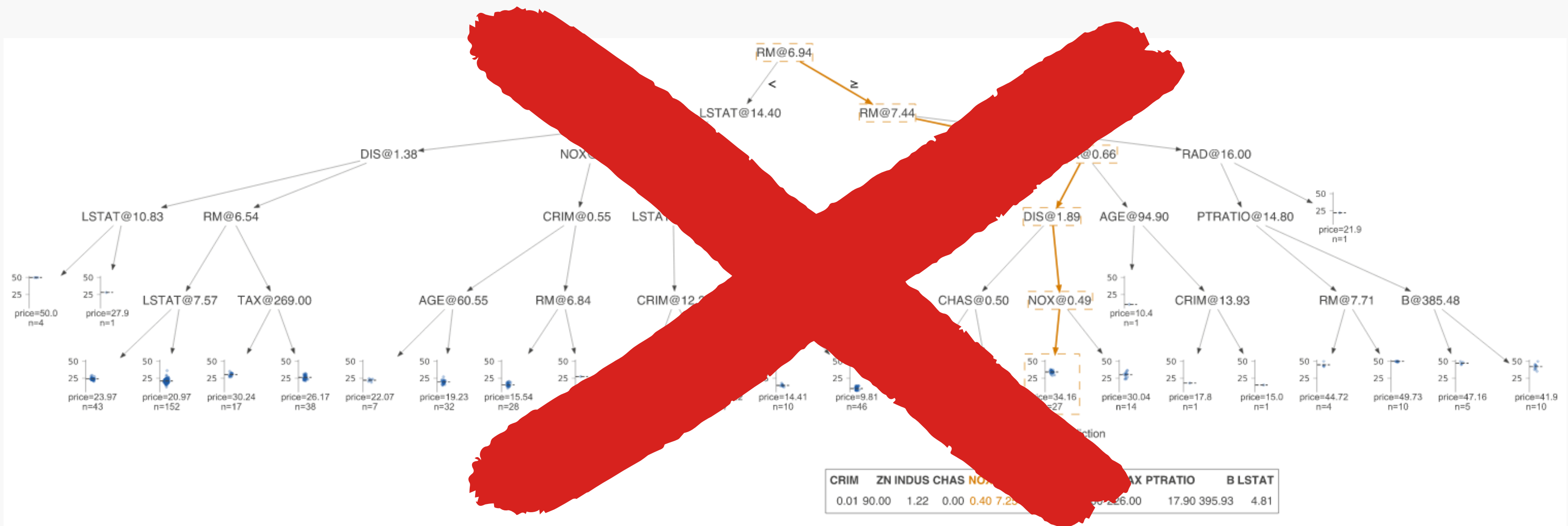


1. Wylosuj obserwacje  
i zbuduj na nich  
pierwsze drzewo  
decyzyjne

2. Utwórz kolejne na  
innym losowym zbiorze  
rows, biorąc pod uwagę  
niedoskonałości  
poprzedniego modelu

3. Powtarzaj krok drugi :)





# Najlepszy gdy...



- jest na czym trenować! +1000 obserwacji
- liczba atrybutów dużo mniejsza niż liczba wierszy (czyli standardowo)
- przeważają wartości numeryczne

# Trochę gorszy kiedy...



- klasyfikujemy grafiki
- rozpoznajemy obrazy (computer vision)
- przetwarzamy język naturalny
- zbiór treningowy taki nie za duży