

# Modelowanie Statystyczne w Zarządzaniu Wierzytelnościami Masowymi

## Laboratorium 4.

1. Wykonaj analizę eksploracyjną danych z poprzedniej listy (p. 2-5).
2. Do zbioru cech aplikacyjnych dodaj 6-cio miesięczną skuteczność (skuteczność to suma wpłat podzielona przez sumę zadłużenia początkowego).
3. Sprawdź oszacowanie treningowego błędu prognozy (zbiór uczący to jednocześnie zbiór testowy) regresyjnego modelu k-najbliższych sąsiadów prognozującego 6-cio miesięczną skuteczność za pomocą cech aplikacyjnych:
  - wykorzystaj funkcję `knn.reg` z pakietu `FNN`
  - sprawdź dla liczb najbliższych sąsiadów od 1 do 30 oraz dla wąskiego oraz szerokiego zbioru cech objaśniających
  - w tym i kolejnych zadaniach jako funkcję straty przyjmij funkcję kwadratową:

$$L(Y, \hat{f}(X)) = (Y - \hat{f}(X))^2$$

- jakie są problemy związane z bezpośrednim prognozowaniem skuteczności?
  - jaka jest optymalna liczba cech objaśniających oraz liczba najbliższych sąsiadów?
4. Sprawdź oszacowanie testowego błędu prognozy regresyjnego modelu k-najbliższych sąsiadów prognozującego 6-cio miesięczną skuteczność za pomocą cech aplikacyjnych:
    - wykorzystaj próbę treningową, walidacyjną oraz testową w proporcjach 50%, 25%, 25%
    - załóż stały zbiór cech objaśniających wybrany w poprzednim zadaniu
    - jak zachowuje się oszacowanie błędu testowego w zależności od liczby najbliższych sąsiadów?
    - jaka jest optymalna liczba najbliższych sąsiadów i związane z nią oszacowanie testowego błędu prognozy?
  5. Sprawdź oszacowanie testowego błędu prognozy regresyjnego modelu k-najbliższych sąsiadów prognozującego 6-cio miesięczną skuteczność za pomocą cech aplikacyjnych wykorzystując 5 i 10-krotną krosvalidację. Sprawdź dla liczb najbliższych sąsiadów od 1 do 30 oraz porównaj z wynikami z poprzednich punktów.
  6. Sprawdź oszacowanie testowego błędu prognozy regresyjnego modelu k-najbliższych sąsiadów prognozującego 6-cio miesięczną skuteczność za pomocą cech aplikacyjnych wykorzystując metodę bootstrap oraz leave one-out bootstrap dla optymalnej liczby najbliższych sąsiadów wybranej w poprzednim punkcie.
  7. Zbierz oraz porównaj wyniki uzyskane w punktach (3-6).