

# Modelowanie Statystyczne w Zarządzaniu Wierzytelnościami Masowymi

## Laboratorium 11.

Na podstawie zbioru danych aplikacyjnych oraz behawioralnych dostępnych w pliku `KrukUWr2018.RData` wykonaj poniższe polecenia.

1. Przygotuj dane aplikacyjne do analizy.
2. Zbuduj na próbie uczącej oraz przetestuj na próbie testowej uogólniony model addytywny prognozujący poziom wpłat w horyzoncie 12 miesięcy obsługi w dwóch podejściach:
  - ① Bezpośrednia prognoza SR12M dla spraw w modelu z ciągłą cechą objaśnianą (wybierz funkcję wiążącą oraz zapisz postać modelu)
  - ② Przełożenie SR12M w przedziałach scorów uzyskanych w modelu z binarną cechą objaśnianą, określającą dobroć klienta (wybierz cechę objaśnianą i funkcję wiążącą oraz zapisz postać modelu)
    - a. W każdym z modeli wytypuj cechy - kandydatki do nieparametrycznej reprezentacji.
    - b. Wykonaj automatyczny dobór cech za pomocą funkcji *step.gam* oraz wybierz najlepszą Twoim zdaniem specyfikację modelu.
    - c. Oceń statystyczną istotność modeli.
    - d. Dla każdego modelu przeanalizuj wykresy cząstkowej predykcji (*partial prediction plots*) – które cechy najsilniej wpływają na dobroć klientów?
    - e. W przypadku modelu logitowego narysuj wykres krzywej ROC i zinterpretuj bazujące na nim miary jakości modelu.
    - f. Który model jest lepszy ze statystycznego punktu widzenia (*explained deviance*, *residual sum of squares*, *AIC* lub inne), a który lepiej prognozuje wpłaty?
    - g. W każdym z podejść sprawdź czy nie wystąpiło zjawisko „współkrzywoliniowości” (*concurvity*).