

# *AutoML*

prof. UAM dr hab. Tomasz Górecki

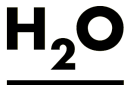
`tomasz.gorecki@amu.edu.pl`

Uniwersytet im. Adama Mickiewicza w Poznaniu  
Wydział Matematyki i Informatyki



The logo for H2O.ai, featuring the text "H2O.ai" in a bold, sans-serif font. The "H2O" is black and the ".ai" is a lighter grey. The logo is centered within a solid yellow square.

**H2O** to darmowa platforma analityczna. W przeciwieństwie do tradycyjnych narzędzi analitycznych H2O zapewnia kombinację najlepszych algorytmów z przetwarzaniem równoległym z łatwością wykorzystania i wdrożenia. H2O oferuje wsparcie dla R, Python, Scala oraz Java. Od 2016 można zaobserwować ogromny wzrost zainteresowania i popularności H2O która nabiera na sile w ostatnim czasie z uwagi na sukcesy jej użytkowników w konkursach publikowanych na **Kaggle**.

The H2O logo, consisting of the text "H2O" in a large, bold, black sans-serif font. A thick horizontal line is positioned directly beneath the "O".

- 1 Głębokie sieci neuronowe (*ang. multi-layer feedforward ANN trained with stochastic gradient descent with backpropagation*)
- 2 Naiwny Bayes
- 3 Generalized Linear Models (GLM)
- 4 Gradient Boosting Machine (GBM)
- 5 Extreme Gradient Boosting (XGBoost)
- 6 Distributed Random Forest (DRF) i Extremely Randomized Trees (XRT)
- 7 Stacked ensembles (klasyfikacja i regresja)
- 8 Model proporcjonalnych hazardów Coxa (analiza przeżycia)
- 9 PCA (redukcja wymiaru)
- 10 Metoda k-średnich (analiza skupień)
- 11 Lasy izolacyjne (wykrywanie anomalii)

**AutoML** jest funkcją H2O, która automatyzuje proces budowania dużej liczby modeli. Celem jest znalezienie najlepszego modelu bez żadnej wiedzy wstępnej. Istnieją implementacje w R i Python. Obecna wersja AutoML wykonuje:

- 1 Trenuje i waliduje: Random Forest, Extremely Randomized Forest, Gradient Boosting Machines (GBMs), Deep Neural Nets i GLM.
- 2 AutoML następnie trenuje dwa modele łączone (*ang. Stacked Ensemble*).
- 3 Pierwszy model łączony zawiera wszystkie uczone modele, drugi zawiera tylko najlepsze z każdej klasy modeli.

# MNIST – klasyfikacja cyfr

MNIST jest dobrze znanym zbiorem benchmarkowym używanym w klasyfikacji. Zbiór uczący składa się z 66000 obrazków, a testowy z 10000 obrazków. Każdy obrazek składa się z  $28^2 = 784$  pikseli w odcieniach szarości.

