

# Projekt z ADD: Filtrowanie reklam

S. Hoa Nguyen

## 1 Opis zadania

Dane prezentują zbiór możliwych reklam na stronach internetowych. Cechy kodują położenia obrazu, frazy występują na adresie URL, tekst kotwicy, słowa występujące obok kotwicy. Zadaniem jest predykcja, czy strona zawiera reklamę (ad"), czy nie (nonad").

## 2 Zadanie

### Selekcja cech:

Ze względu na to, że dane zawiera ponad 1500 cech, zatem koniecznym krokiem jest usunięcie cech nieistotnych.

### Budowanie modelu:

Budowanie model predykcyjny do rozpoznawania reklam internetowych.

## 3 Model testowania

- a) Zbiór treningowy: (70% przykładów) **trn**
- b) Zbiór testowy: (30% przykładów) **tst**
- c) Miara oceny: Accuracy.

## 4 Wymaganie

- a) Testować klasyfikator obliczony na danych oryginalnych.
- b) Testować klasyfikator obliczony na danych po selekcji.
- c) Porównywać wyniki, napisać swoje obserwacje i wnioski.
- d) Czy selekcja cech zwiększa skuteczność klasyfikacji?

## 5 Źródło danych

<https://archive.ics.uci.edu/ml/datasets/Internet+Advertisements>