

# Sztuczna inteligencja

## Laboratorium

### Lista 1

termin: pierwsze zajęcia w listopadzie

Do rozwiązania zadań należy wykorzystać język programowania Python wraz z pakietem scikit-learn. Przed zainstalowaniem scikit-learn należy zainstalować pakiety NumPy oraz SciPy. Przydatne będą również pakiety Matplotlib oraz Pandas. Ta strona (choć nieoficjalna) może okazać się bardzo pomocna.

**Zad. 1** — Zapoznaj się z podstawowym tutorialiem pakietu scikit-learn. Wczytaj standardową bazę danych *iris*. Przyjrzyj się danym, zrozum, co opisują.

1. Wyświetl podstawowe statystyki każdej z czterech cech (liczność, średnią, odchylenie standardowe, minimum, maksimum i kwartyle). Sprawdź, ile instancji obiektów znajduje się w każdej klasie. [*describe()*, *groupby()*]
2. Dla każdej z czterech cech wyświetl wykres pudełko i wąsy oraz histogram (zrozum, co przedstawiają).

**Zad. 2** — Podziel dane na zbiór danych treningowych (80%) oraz testowych (20%). Po co to się robi? [*cross\_validation.train\_test\_split()*]

Przetestuj trzy klasyfikatory: regresję logistyczną, drzewa decyzyjne oraz metodę  $k$  najbliższych sąsiadów (naucz się podstaw każdej z nich). Porównaj, jak wypadły na zbiorze testowym, obliczając ich błędy realne (załóż, że dane pochodzą z rozkładu jednostajnego).