

# Wstęp do sztucznej inteligencji

## Laboratorium 1

3 marca 2023

### Zasady oceniania

W ramach zajęć należy po kolei przerobić zadania umieszczone w dalszym ciągu tej instrukcji. W wyniku pracy ma powstać raport w formacie PDF, zawierający nastawienia metod, wyniki w postaci wykresów i macierzy konfuzji, a także wnioski. Dodatkowo, należy przesłać kod powstały w ramach zajęć. Akceptowalny jest też wydruk z Colab/Jupyter Notebook zapisany **w formacie pdf**.

Raport z wykonanych zadań **w formacie PDF** proszę przesłać na platformie Moodle wraz z kodami źródłowymi programów (pliki **.py**).

**UWAGA:** Termin oddania zadania jest ustawiony w systemie moodle. W przypadku nie oddania zadania w terminie, uzyskana ocena będzie zmniejszana o 0,5 za każdy zaczęty tydzień opóźnienia. Zadania oddawane później niż miesiąc po terminie ustawionym na moodle są oddawane i rozliczane w trybie indywidualnym na zajęciach lub po umówieniu się z prowadzącym.

**UWAGA:** W przypadku wysłania zadania w formie niezgodnej z opisem w instrukcji prowadzący zastrzega prawo do wystawienia oceny negatywnej za taką pracę. Przykład: wysłanie **.zip** lub **.pdf** tam, gdzie był wymagany plik tekstowy z rozszerzeniem **.py**.

### 1 Zadania do wykonania

1. Zainstaluj biblioteki `numpy`, `matplotlib` i `scikit-learn`.
2. Zapoznaj się z funkcjonalnościami dostępnymi w podmodule `sklearn.datasets` <https://scikit-learn.org/stable/datasets.html#datasets>. Zadania w tym laboratorium będą wykonywane na zbiorze danych Iris.
3. Zapoznaj się z dokumentacją klasy `KNeighborsClassifier` oraz opisem metody najbliższych sąsiadów.  
<https://scikit-learn.org/stable/modules/neighbors.html#classification>.  
<https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>
4. Dla zbioru danych Iris wykonaj:  
Iris: [https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load\\_iris.html#sklearn.datasets.load\\_iris](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_iris.html#sklearn.datasets.load_iris)
  - (a) Wygeneruj/wczytaj zbiór danych. W przypadku Iris będą to cztery zmienne (kolumny), i 150 próbek (wierszy).
  - (b) Podziel dane na zbiór uczący i testujący za pomocą funkcji `train_test_split` w proporcji 0.7 danych uczących i 0.3 testowych: [https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.train\\_test\\_split.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html).
  - (c) Naucz klasyfikator `KNeighborsClassifier` na danych uczących zadaną liczbą sąsiadzi  $N$ .
  - (d) Oblicz dokładność klasyfikacji (accuracy) na zbiorze danych testowych. [https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy\\_score.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html).
  - (e) Powtórz punkty 4c i 4d w celu znalezienia optymalnej liczby najbliższych sąsiadzi (takiej przy której by była uzyskana największa dokładność). Można wykonywać poprzednie punkty w pętli i narysować wykres pokazujący zależność dokładności od liczby sąsiadzi  $N$ .

- (f) Dla wybranej wartości  $N$ , wykonaj:
- Naucz klasyfikator na danych uczących ostatecznie dla wybranej wartości  $N$ .
  - Stwórz macierz konfuzji, pokazującą dokładność klasyfikacji na danych testowych ([https://scikit-learn.org/stable/auto\\_examples/model\\_selection/plot\\_confusion\\_matrix.html#confusion-matrix](https://scikit-learn.org/stable/auto_examples/model_selection/plot_confusion_matrix.html#confusion-matrix)).
  - Wykonaj wizualizację całości danych za pomocą wykresu punktowego (scatter), pokazującego rzeczywisty podział klas, a następnie wykonaj taki sam wykres, ale zamiast rzeczywistego wektora klas użyj wynik predykcji zbudowanego klasyfikatora. Punkty na wykresie muszą być pokolorowane według klas!  
W przypadku zbioru Iris wizualizację należy wykonać za pomocą dwóch pierwszych zmiennych (sepal length, sepal width).
  - Dodatkowo dla zbioru Iris należy wykonać wizualizację za pomocą trójwymiarowego wykresu punktowego, na podstawie trzech zmiennych (sepal length, sepal width, petal length). Punkty muszą być pokolorowane według klas.
5. W raporcie proszę umieścić wyniki poszczególnych etapów wykonywania zadania, jak to zrzuty ekranu części wygenerowanych/wczytanych danych, dobrana liczba  $N$  sąsiedzi, ewentualnie wyniki dokładności dla innych liczb  $N$ , wykresy pokazujące rzeczywisty podział klas oraz podział klas według predykcji nauczonego klasyfikatora.