

# Kolokwium zaliczeniowe z przedmiotu Podstawy Sztucznej Inteligencji

Mgr inż. Jolanta Podolszańska

## 1 Ogólne informacje

Podczas kolokwium będą Państwo projektować program w języku Python w oparciu o zbiór danych, który znajdzie Państwo w sekcji kolokwium. Proszę o uważne czytanie pytań. Używanie **Chat GPT** lub innych chatbotów/aplikacji opartych o AI do generowania rozwiązań jest **zabronione**, tak samo jak korzystanie z telefonu podczas kolokwium. Można korzystać ze sporządzonych notatek na kolokwium, o których mówiłam na zajęciach. Z Internetu korzystamy tylko i wyłącznie do przesłania rozwiązanych kolokwium.

Maksymalna liczba punktów do zdobycia to 12 punktów. Zaliczenie kolokwium jest od 6 pkt.

## 2 Zadania

Proszę o wykonanie następujących zadań.

1. Wczytaj plik *heart.csv* do ramki DataFrame z biblioteki *Pandas*. Wyświetl pierwsze wiersze (`df.head()`) i nazwy kolumn (`df.columns`). Sprawdź, ile wierszy (próbek) i ile kolumn (cech + etykieta) posiada zbiór. Zwróć uwagę, czy w pliku występują kolumny zbędne, np. zawierające same wartości NaN. Jeśli tak, zdecyduj, co z nimi zrobić (usuń je przy pomocy `df.drop`) (2 pkt)

Przykład usuwania kolumn

```
\begin{lstlisting}
if 'id' in df.columns:
    df.drop('id', axis=1, inplace=True)
if 'Unnamed: 32' in df.columns:
    df.drop('Unnamed: 32', axis=1, inplace=True)
```

2. Zidentyfikuj kolumnę z etykietą – zwykle target, w której 0 oznacza brak choroby, a 1 oznacza chorobę serca. Wybierz tylko dwie kolumny (np. *chol* – cholesterol, *trestbps* – ciśnienie krwi w spoczynku) jako cechy. Zapisz je w zmiennej *X\_2d*. Wektor etykiet zapisz w zmiennej *y* (2 pkt).

3. Podziel zbiór danych na zbiór treningowy, testowy i walidacyjny. Najpierw rozdziel na zbiór treningowy (np. 70%) i tymczasowy (30%), Następnie ten tymczasowy 30% podziel na walidację (50% z 30% = 15% całości) i test (15%). Wypisz rozmiary uzyskanych zbiorów – upewnij się, że sumują się do całej liczby próbek (3 pkt ).
4. Wykonaj standaryzację cech (StandardScaler) na zbiorze treningowym (fit\_transform) oraz na zbiorze walidacyjnym i testowym (transform). Wytrenuj perceptron (np. Perceptron(max\_iter=5000, eta0=0.001, random\_state=42)) na zbiorze treningowym (2 pkt).
5. Połącz wszystkie próbki (train + val + test) do jednego zbioru, aby je narysować na wspólnym wykresie. Podziel punkty na klasę 0 (brak choroby) i klasę 1 (choroba), nadaj im różne kolory (np. czerwony / niebieski). Oblicz i narysuj linię decyzyjną perceptronu:

$$y = \frac{b + w_1 x}{w_2}$$

gdzie  $w_1, w_2 = \text{perceptron.coef\_}[0]$  i  $b = \text{perceptron.intercept\_}[0]$ . Użyj do tego biblioteki *Matplotlib* (3 pkt).