

---

# METODY SYSTEMOWE I DECYZYJNE L

## RAPORT ZALICZENIOWY

---

Damian Serwata  
W4N, K46  
damian.serwata@pwr.edu.pl

sem. letni 2023/24

## 1 Zadanie

### 1.1 Treść

Celem zadania zaliczeniowego jest przeprowadzenie analizy wybranego zbioru danych bazując na wiedzy i umiejętnościach uzyskanych podczas kursu <sup>1</sup>.

Przykładowo możemy zaimplementować model dobierający obroty wentylatora w komputerze na podstawie poziomu naładowania baterii, informacji o tym czy jest ładowana, jaki procent CPU oraz GPU jest używany. Możemy także spróbować przewidywać kto przeżył katastrofę statku RMS Titanic biorąc pod uwagę płeć, wiek, klasę biletu. Wreszcie innym przykładem jest przewidywanie liczby zachorowań na COVID-19 na podstawie danych pogodowych oraz mobilności populacji w danym mieście.

**Nie jest dopuszczalne wykorzystanie tych samych danych przez kilku studentów.** Dodatkowo wysoce rekomenduje się pracę na samodzielnie pozyskanie danych.

Za rozwiązanie zadania uznaje się raport opisujący pokonywany problem, dostępne dane, przyjęte metody (wraz z uzasadnieniem), krok po kroku wykonaną pracę implementacyjną oraz **wnioski**.

### 1.2 Artefakty

Po wykonaniu zadania proszę o wysłanie na ePortalu następujących artefaktów:

- raport w formacie PDF,
- pliki źródłowe  $\LaTeX$ ,
- pliki źródłowe Pythona,
- dane będące przedmiotem analizy.

W przypadku gdy rozmiar analizowanych danych przekracza ten dopuszczalny na ePortalu, proszę o przesłanie na ePortalu reprezentatywnego wycinka danych oraz udostępnienia reszty w inny sposób - np. poprzez Google Drive.

Pliki powinny być spakowane do archiwum nazwanego w postaci: "*raport\_{imię}\_{nazwisko}\_{numer\_grupy}.zip*", np. "*raport\_jan\_kowalski\_3.zip*". Proszę zadbać aby było możliwe ponowne uruchomienie skryptów i kompilacja raportu przeze mnie.

---

<sup>1</sup>oczywiście można próbować wykraczać poza materiał kursu z zastrzeżeniem, że autor rozumie wykorzystywane narzędzia

## 2 Punktacja i warunki zaliczenia

### 2.1 Laboratorium

Wskaźnik  $W$ , który służy do określenia oceny końcowej, obliczamy w następujący sposób:

$$W = (R + A) \times F$$

gdzie

$R$  - liczba punktów z raportu zaliczeniowego,  $R \in [0, 10]$

$A$  - liczba punktów z aktywności podczas zajęć,  $A \in [0, 3.5]$

$F$  - wskaźnik obecności (frekwencji),  $F = \begin{cases} 1, & \text{jeśli liczba nieobecności} \leq 1 \\ 0, & \text{w innych przypadkach} \end{cases}$

Kończącą ocenę określamy w następujący sposób:

Ocena	Liczba punktów $W$
2	$\leq 6.5$
3	$(6.5, 7.5]$
3.5	$(7.5, 9]$
4	$(9, 10.5]$
4.5	$(10.5, 12]$
5	$(12, 13]$
5.5	$> 13$

### 2.2 Zadanie zaliczeniowe

Wskaźnik  $R$ , będący oceną za raport końcowy, określamy w następujący sposób:

$$R = d + p + s + e + r$$

gdzie

$d$  - dobór danych i ich analiza,  $d \in [0, 3]$

$p$  - rozwiązanie problemu,  $p \in [0, 4]$

$s$  - styl kodu w języku Python,  $s \in [0, 1]$

$e$  - estetyka raportu  $e \in [0, 1]$

$r$  - regularność raportowania postępu prac  $r \in [0, 1]$

#### Dobór danych i ich analiza

Ocenie podlegać będzie sposób zgromadzenia zbioru (sztuczny lub powszechnie dostępny będzie oceniany mniej korzystnie niż samodzielnie pozyskane dane), opis jego kontekstu oraz wykorzystane metod pre-processingu oraz oczyszczania danych. Wreszcie wstępna analiza (tzw. exploratory data analysis)<sup>2</sup> w celu określenia zależności między danymi, odrzucenia danych redundantnych, outlierów itp. Maksymalna możliwa ocena tej części, w przypadku gotowych zbiorów (nie pozyskanych przez studenta), wynosi 1 punkt.

#### Rozwiązanie problemu

Ocenie podlegać będzie sam problem rozwiązywany przez autora, a także sposób jego rozwiązania. Przede wszystkim należy zastanowić się nad sensem wyzwania (np. czy jest sens badać związek nasłonecznienia z ceną benzyny), a także nad jasnym sformułowaniem rozwiązywanego problemu.

Przedstawiony problem powinien być rozwiązany używając metod klasyfikacji bądź regresji. Należy zastanowić się nad doбором, zrozumieniem i uzasadnieniem użytych metod. Wskazane jest także porównanie działania modelu (modeli) na różnych parametrach.

<sup>2</sup>do tego można użyć m.in. biblioteki pandas-profiling

### **Styl kodu w języku Python**

Ocenie podlegać będzie "inżynierskie podejście" do tworzenia oprogramowania, tj. styl kodu (zgodność z PEP8)<sup>3</sup>, dodanie komentarzy, brak polskich zmiennych, modularność kodu oraz zapewnienie jego reprodukowalności, itp...

### **Estetyka raportu**

Ocenie podlegać będzie przejrzystość raportu, poprawność i styl językowy, jakość wizualizacji danych i wyników oraz ich opisów.

### **Regularność raportowania postępu prac**

Oceniane będzie również regularne raportowanie postępu prac poprzez zwięzłe raporty podczas rozmowy z prowadzącym w ramach zajęć.

## **3 Terminarz**

Zadanie proszę wysłać w terminie do dwóch dni poprzedzających ostatnie zajęcia, tj.:

- 23:59:59 CEST, 20.05.23 (grupa 5),
- 23:59:59 CEST, 10.06.23 (grupy 6).

Na przedostatniach laboratoriach przedstawię Państwu proponowaną ocenę.

---

<sup>3</sup>tu można użyć linterów, np. black lub flake8