

Home page

Dydaktyka – bieżące

- [AiR – Inf1](#)
- [AiR – PAOM](#)
- [AiR – ML](#)
- [Inf – AdvML](#)

- [Prace mgr 2018/2019](#)

Przydatne

- [Jak korzystać z wiki?](#)
- [WiFi AGH](#)
- [UCI AGH](#)

# Wprowadzenie do uczenia maszynowego. Podstawy Python

## Slajdy


Sprawy organizacyjne:  [Slajdy](#)

Wprowadzenie:  [Intro](#)

## Podstawy Python - przetwarzanie danych

Jednym z podstawowych pakietów do analizy danych w języku Python jest pakiet Pandas. W trakcie wykonywania zadań proszę wykorzystać podstawowe operacje związane z analizą danych za pomocą funkcji dostępnych w pakiecie Pandas.

Proszę zapoznać się z  [Pandas Cheat Sheet](#)

Proszę przeanalizować przykłady  [10 minutes - Panda](#)

Pierwszym krokiem do rozpoczęcia pracy z Python i Pandas jest zaimportowanie niezbędnych pakietów. W proponowanym przykładzie należy zaimportować pakiet NumPy do obliczeń numerycznych, zwłaszcza obliczeń wektorowo-macierzowych, pakiet Matplotlib do tworzenia wykresów oraz oczywiście sam pakiet Pandas. Wymienione pakiety zostały domyślnie umieszczone w popularnej dystrybucji Anaconda Python używanej w Data Science i są domyślnie instalowane

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

### Zadanie 1 - baza danych Iris

Proszę zapoznać się z bazą danych Iris  [Iris flowers](#)

estaw pomiarów kwiatów irysa, udostępniony po raz pierwszy przez Ronalda Fishera w roku 1936. Jeden z najbardziej znanych zbiorów, a zarazem bardzo prosty i użyteczny. Zbiór irysów składa się z 4 wartości pomiarów jego płatków (szerokości i długość) oraz klasy do jakiej należy. Przykład kilku wybranych rekordów:

```
sepal length,sepal width,petal length,petal width,class
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
7.0,3.2,4.7,1.4,Iris-versicolor
```

Import danych:

```
from sklearn.datasets import load_iris
# load the famous iris data
irisRaw = load_iris()

#Konwersja danych do pakietu Panda
# read iris.data into a pandas DataFrame (df), including column names
iris = pd.DataFrame(data= np.c_[irisRaw['data'], irisRaw['target']],
                    columns= irisRaw['feature_names'] + ['target'])
```

- Proszę wyświetlić dane i zapoznać się z poszczególnymi informacjami.
- Wyświetl liczbę wierszy oraz kolumn.
- Wyświetl podstawowe informacje dla poszczególnych kolumn. Skorzystaj z metod `describe` oraz `groupby`
- Wykorzystując metodę `head` proszę wyświetlić 5 pierwszych wierszy
- Korzystając z `dropna` sprawdź, czy baza zawiera brakujące dane
- Sortowanie zbioru danych według określonych kryteriów to kolejny ważny element pakietu pandas. W celu wykonania sortowania po wierszach lub kolumnach należy skorzystać z metody `sort_index`, która zwraca nowy, posortowany obiekt. Posortuj dane rosnąco względem drugiej kolumny.
- Wyznacz minimalną i maksymalną długość płatka (kolumna 3 - petal length). Podaj indeks tych wartości.
- Oblicz odchylenie standardowe dla każdej kolumny
- Wyodrębnij kolumny dla których długość kielicha kwiatów (sepal length) jest większa od średniej długości (cały zbiór)

Proszę zapoznać się z możliwościami generowaniem wykresów:

 [PyPlot tutorial](#)

- Wyświetl histogramy dla poszczególnych parametrów z uwzględnieniem przynależności do danej klasy

teaching:air-ml:2019I:labs:ab01

#### Table of Contents

- ◊ [Wprowadzenie do uczenia maszynowego. Podstawy Python](#)
- ◊ [Slajdy](#)
- ◊ [Podstawy Python - przetwarzanie danych](#)
  - ◊ [Zadanie 1 - baza danych Iris](#)

