

Biblioteki Python o określonej funkcjonalności

Python to język programowania, który jest popularny m.in. dzięki dużej liczbie dostępnych bibliotek. Dzięki nim można łatwo i szybko korzystać z różnych funkcji – na przykład do analizy danych, tworzenia grafiki czy automatyzacji zadań.

Wybrane biblioteki:

Analiza danych

- **Pandas** – popularna biblioteka do pracy z danymi w tabelach (DataFrame). Umożliwia łatwe filtrowanie, grupowanie i obliczenia.
- **Polars** – nowoczesna i bardzo szybka alternatywa dla Pandas. Działa wydajnie nawet przy dużych zbiorach danych.
- **Dask** – świetna do pracy z bardzo dużymi plikami danych. Dzieli dane na mniejsze części i przetwarza je równolegle.

Wizualizacja danych

- **Matplotlib** – podstawowe narzędzie do rysowania wykresów 2D, np. liniowych czy słupkowych.
- **Seaborn** – ułatwia tworzenie estetycznych wykresów statystycznych, bazując na Matplotlib.
- **Plotly** – pozwala tworzyć nowoczesne, interaktywne wykresy, które można przeglądać w przeglądarce (np. w dashboardach).

Przetwarzanie obrazów

- **Pillow** – prosta biblioteka do podstawowych operacji na obrazach, np. rozmycia, kadrowania, zapisywania.
- **OpenCV** – zaawansowane narzędzie do przetwarzania obrazu, wykrywania krawędzi i obiektów.
- **scikit-image** – biblioteka naukowa, która oferuje różne filtry i metody analizy obrazów.

Web scraping

- **Selenium** – narzędzie, które pozwala automatycznie obsługiwać przeglądarkę, np. w celu pobrania nagłówków ze stron internetowych (jak Onet.pl).

Dla Selenium wykorzystano wbudowaną przeglądarkę Safari (macOS) lub Chrome (inne systemy). W przykładach zastosowano rzeczywiste pliki (np. data.csv, example.jpg).

Przykłady użycia

Analiza danych (plik: data_analysis.py)

Przy użyciu Pandas, Polars i Dask wykonano statystyki opisowe na zbiorze danych.

```
Pandas describe():
              33863          4.0
count    49.000000  49.000000
mean    33479.000000  3.602041
std       290.178465  0.835287
min     33022.000000  2.000000
25%     33262.000000  3.500000
50%     33425.000000  4.000000
75%     33763.000000  4.000000
max     33997.000000  5.000000
```

```
Polars describe():
shape: (9, 5)
```

statistic	Emilia	Kaczmarek	33863	4.0
---	---	---	---	---
str	str	str	str	str
<hr/>				
count	49	49	49	49
null_count	0	0	0	0
mean	null	null	null	null
std	null	null	null	null
min	Adam	Grabowska	33022	2.0
25%	null	null	null	null
50%	null	null	null	null
75%	null	null	null	null
max	Zofia	Wiśniewski	33997	5.0

```
Dask mean (tylko kolumny numeryczne):
```

```
33863    33479.000000
```

```
4.0      3.602041
```

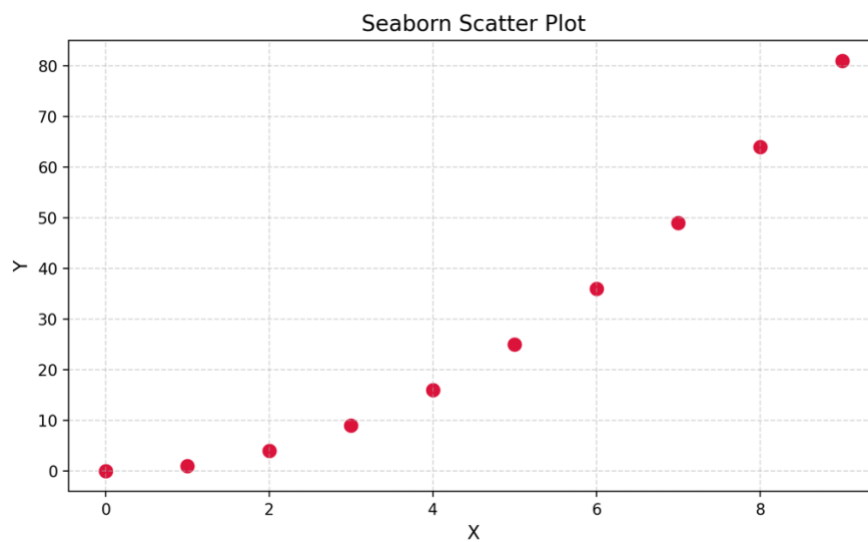
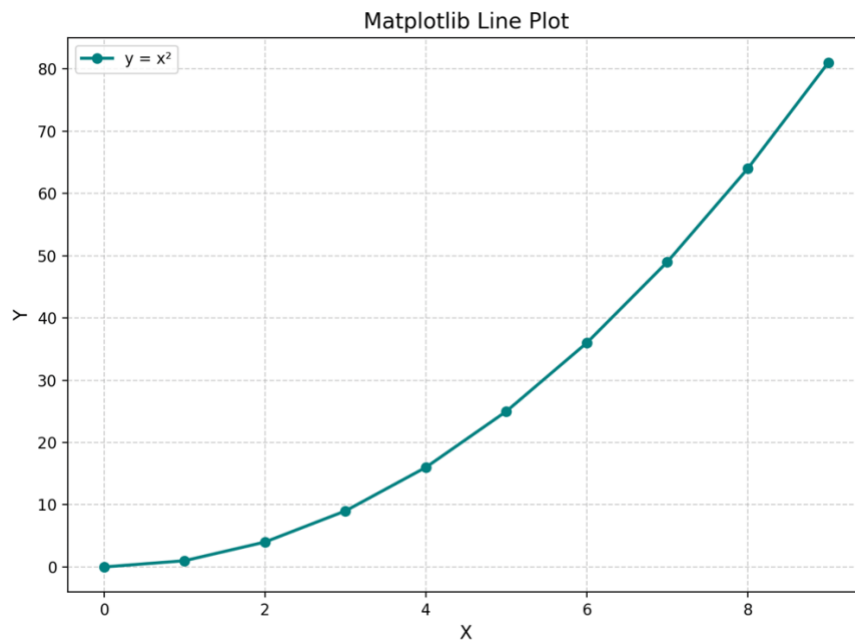
```
dtype: float64
```

```
Process finished with exit code 0
```

📊 Wizualizacja danych (plik: visualization.py)

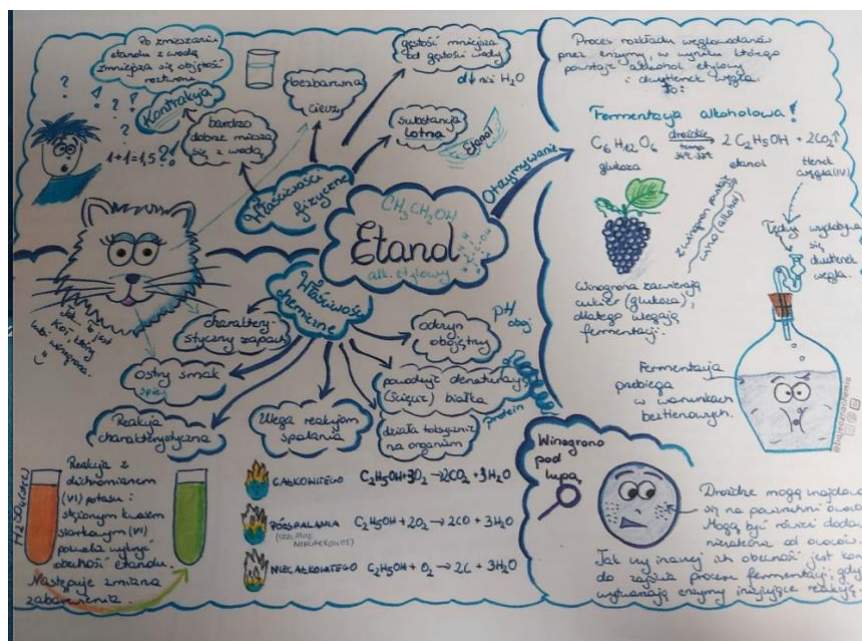
Stworzono wykresy dla funkcji kwadratowej:

- Liniowy wykres z Matplotlib
- Scatter plot z Seaborn
- Interaktywny wykres z Plotly



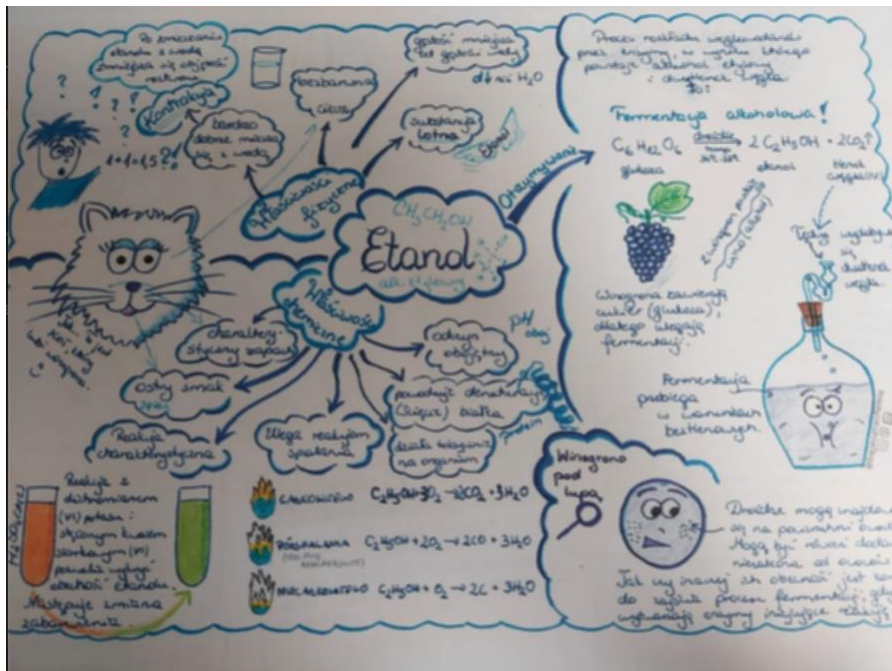
- Pillow: zastosowano rozmycie na zdjęciu
- OpenCV: wykryto krawędzie (Canny)
- scikit-image: użyto filtra Sobela

ORYGINAŁ:

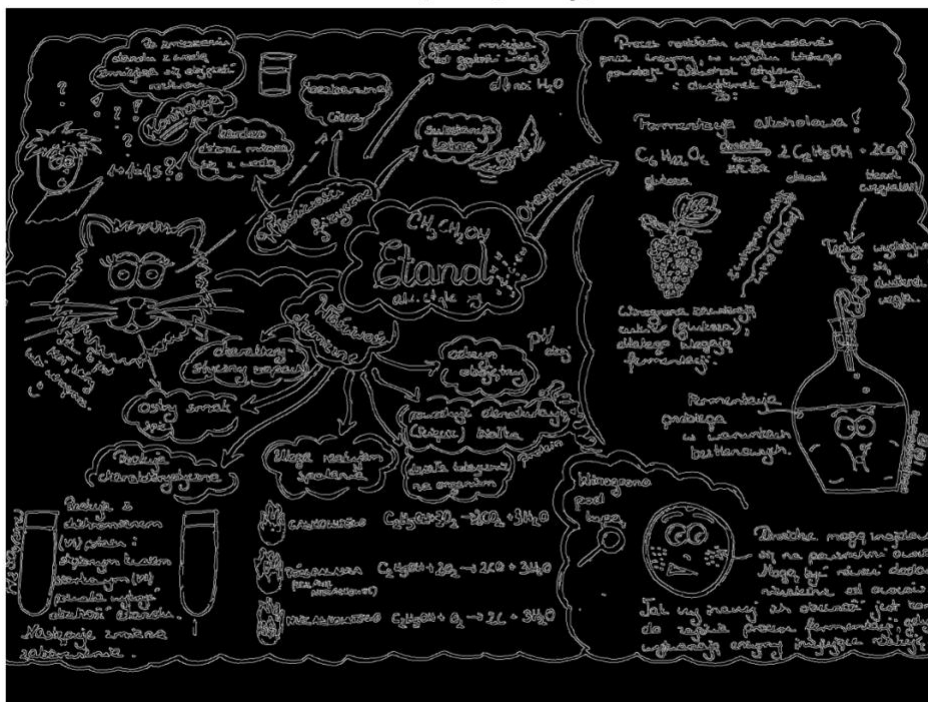


(zdjęcie: @bajecznachemia)

PO ZMIANACH:



Krawędzie (Canny)



Filtr Sobela (scikit-image)



🌐 Web scraping (plik: web_scraping.py)

Selenium automatycznie otwiera przeglądarkę i pobiera nagłówki ze strony Onet.pl.

```
"/Users/monikaturkowska/Desktop/Biblioteki python/bin/python" /Users/monikaturkowska/Desktop/Biblioteki python/web_scraping.py
System iOS wykryty - uruchamiamy Safari.
Otwieram Onet...

Nagłówki z Onet.pl:
- O tym się mówi
- NBP poszedł pod prąd. Złoto okazało się strzałem w dziesiątkę
- Oficer KGB był wstrząśnięty prawdą o Katyniu. "Polska przyjęła go jak wroga"
- Potężne burze i wichury wkraczają do Polski. RCB rozeszło ostrzeżenia
- Szokujące słowa Michała Kamińskiego. "Ubeckie metody"
- Media: USA przedstawiły propozycję pokoju między Rosją a Ukrainą
- Nowe fakty w sprawie tragedii na A1. Prokuratura umarza dwa śledztwa
- Serwis Igi Świątek rozsypał się jak domek z kart. Oto powód
- Burza wokół aborcji w Oleśnicy. Jest reakcja RPO
- Co dalej z pokojem w Ukrainie? Kreml reaguje na ostre stanowisko USA

Process finished with exit code 0
```

Pandas

Zalety:

- Łatwy w użyciu.

- Umożliwia szybkie przetwarzanie danych (np. sortowanie, filtrowanie).
- Dużo przykładów i porad w internecie.

Wady:

- Może być wolny przy bardzo dużych plikach.
 - Nie wykorzystuje całej mocy komputera (działa tylko na jednym rdzeniu).
-

Polars

Zalety:

- Bardzo szybki.
- Potrzebuje mniej pamięci niż Pandas.
- Nowoczesny i prosty w użyciu.

Wady:

- Mniej poradników i przykładów.
 - Niektóre funkcje są jeszcze w rozwoju.
-

Dask

Zalety:

- Radzi sobie z bardzo dużymi danymi.
- Może używać wielu rdzeni komputera.
- Podobny do Pandas, więc łatwo się przestawić.

Wady:

- Czasem trudniejszy w użyciu niż Pandas.
 - Przy małych danych działa wolniej.
-

Matplotlib

Zalety:

- Można tworzyć różne wykresy.
- Dużo materiałów i przykładów.
- Można zapisać wykresy do pliku.

Wady:

- Wykresy domyślnie są mniej estetyczne.
 - Trzeba napisać więcej kodu niż w innych bibliotekach.
-

Seaborn**Zalety:**

- Tworzy ładne wykresy automatycznie.
- Łatwy do użycia z danymi statystycznymi.
- Działa razem z Matplotlib.

Wady:

- Mniej opcji przy nietypowych wykresach.
 - Trudniej zmieniać wygląd wykresów.
-

Plotly**Zalety:**

- Wykresy są interaktywne – można klikać, powiększać.
- Świetnie wygląda w przeglądarce.
- Łatwy w użyciu.

Wady:

- Zapisywanie wykresów może być trudniejsze.
 - Więcej plików i zależności w projekcie.
-

Pillow**Zalety:**

- Prosta biblioteka do obrazów.
- Można łatwo przycinać, zmniejszać lub rozmywać zdjęcia.

Wady:

- Tylko do prostych rzeczy.
- Ma mniej funkcji niż OpenCV.

OpenCV

Zalety:

- Bardzo zaawansowana biblioteka do obrazu i wideo.
- Szybka i wydajna.
- Można wykrywać twarze i obiekty.

Wady:

- Trudniejsza w nauce.
- Może być przytłaczająca dla początkujących.

scikit-image

Zalety:

- Dobra do naukowych operacji na obrazach.
- Ma gotowe filtry i funkcje.
- Dobrze działa z innymi bibliotekami.

Wady:

- Wolniejsza niż OpenCV.
- Nie obsługuje wideo.

Selenium

Zalety:

- Może automatycznie obsługiwać strony internetowe.
- Używana do web scrapingu i testów.
- Działa z różnymi przeglądarkami.

Wady:

- Działa wolno (czeka na załadowanie strony).
- Może nie działać, jeśli strona się zmienia.

Linki do dokumentacji:

- Pandas - <https://www.python.org/search/?q=pandas&submit=>
- Polars - <https://pypi.org/project/polars/>
- Dask - <https://www.dask.org>
- Matplotlib - <https://matplotlib.org>
- Seaborn - <https://seaborn.pydata.org>
- Plotly - <https://plotly.com>
- Pillow - <https://pillow.readthedocs.io/en/stable/>
- OpenCV - <https://opencv.org>
- scikit-image - <https://scikit-image.org>
- Selenium - <https://www.selenium.dev>

Podsumowanie

W ramach laboratorium zapoznano się z bibliotekami z czterech różnych dziedzin: analiza danych, wizualizacja, grafika oraz web scraping. Przygotowano działające przykłady i wygenerowano pliki graficzne i tekstowe.