

Przegląd bibliotek i narzędzi do uczenia maszynowego oraz przetwarzanie danych oraz wizualizacja.

Zestaw danych:

https://pwozniak.kia.prz.edu.pl/files/uczeniemaszynowe/train_data.csv

Dane, które analizujemy, pochodzą z pliku CSV, który zawiera następujące kolumny:

- #: Numer Pokemona.
- Name: Nazwa Pokemona.
- Type 1: Główny typ Pokemona.
- Type 2: Dodatkowy typ Pokemona (jeśli dotyczy).
- Total: Łączna suma statystyk (HP, Attack, Defense, Sp. Atk, Sp. Def, Speed).
- HP: Punkty zdrowia.
- Attack: Atak fizyczny.
- Defense: Obrona fizyczna.
- Sp. Atk: Atak specjalny.
- Sp. Def: Obrona specjalna.
- Speed: Szybkość.
- Generation: Generacja Pokemona.
- Legendary: Status legendarności (True/False).

Zadania:

1. Zastosowanie bibliotek do analizy danych (Pandas, NumPy, itd.).
2. Wykrywanie błędnych, brakujących wartości i duplikatów w zbiorze danych.
3. Przetwarzanie istniejących danych.
4. Wizualizacja danych oraz wyznaczenie statystyk.

W projekcie zastosuj następujące kroki:

1. **Wczytanie danych:** Użyjemy Pandas do wczytania pliku CSV.
2. **Eksploracja danych:** Zbadamy podstawowe statystyki i rozkład danych.
3. **Przetwarzanie danych:**
 - Wykrywanie i obsługa brakujących wartości.
 - Usuwanie duplikatów.
 - Sortowanie na podstawie Numeru pokemona.

4. Analiza danych:

- Grupowanie danych według typu Pokemona.
- Analiza liczby wierszy o każdym pokemonie (nazwie).
- Wyznaczenie liczby błędnych danych.
- Wyznaczenie średnich statystyk dla każdego gatunku pokemona.
- Wyznaczenie średnich statystyk dla każdego typu pokemona.
- Analiza statystyk bojowych.

5. Przetwarzanie danych:

- Usuwanie błędnych danych.
- Modyfikowanie istniejących danych.
- Wprowadzanie nowych danych:
<https://gist.github.com/armgilles/194bcff35001e7eb53a2a8b441e8b2c6>
- Zapisywanie zmodyfikowanych danych

6. Wizualizacja danych:

- Utworzenie wykresów pokazujących rozkłady statystyk (matplotlib itp.).
- Wizualizacja zależności między statystykami (np. między atakiem a obroną).
- Wykres liczby pokemonów dla każdego typu głównego.
- Wykres pokemonów z 1 lub 2 typami.
- Wykres t-SNE dla danych 'HP', 'Attack', 'Defense', 'Sp. Atk', 'Sp. Def', 'Speed'

<https://scikit-learn.org/dev/modules/generated/sklearn.manifold.TSNE.html>