**Statystyczne metody rozpoznawania obrazu**

**Zadanie 3**

**Bartłomiej Gorzela**

W celu wykonania zadania posłużyłem się przykładowym zbiorem danych ‘Iris’ możliwym do zaimportowania bezpośrednio z biblioteki scikit-learn dostępnej w Python.

1. **Podać zakres zmienności wartości atrybutów.**

Zakres zmienności wartości atrybutów obliczyłem jako różnicę pomiędzy maksymalną a minimalną z wartości dla każdego z atrybutów.

**Kod źródłowy:**

**A computer code on a black background

AI-generated content may be incorrect.**

**Wynik:**

**A black screen with white text

AI-generated content may be incorrect.**

1. **Obliczyć średnią dla wartości atrybutów wybranego zbioru danych.**

Średnią wartość dla każnego z atrybutów obliczyłem za pomocą dostępnej funkcji mean().

**Kod źródłowy:**

**A black background with white text

AI-generated content may be incorrect.**

**Wynik:**

A screenshot of a computer

AI-generated content may be incorrect.

1. **Obliczyć odchylenie standardowe wartości atrybutów.**

Odchylenie standardowe wartości każego z atrybutów policzyłem za pomocą dostępnej funkcji std().

**Kod źródłowy:**

**A black screen with white text

AI-generated content may be incorrect.**

**Wynik:**

**A screenshot of a computer

AI-generated content may be incorrect.**

1. **Przeskalować do pewnego przedziału. Wybraną metodą zwizualizować różnice między tak otrzymanymi wartościami cech a tymi przed przeskalowaniem.**

Wartości każdego z atrybutów przeskalowałem do przedziału [0, 1] przy pomocy obiektu MinMaxScaler. Różnicę przed i po skalowaniu przedstawiłem za pomocą histogramów.

**Kod źródłowy:**

**A screen shot of a computer program

AI-generated content may be incorrect.**

**Wynik:**

A group of graphs showing different sizes of bars

AI-generated content may be incorrect.

1. **Wykonać standaryzację wartości atrybutów. Wybraną metodą zwizualizować różnice.**

Standaryzacje dla każdego z atrybutów wykonałem przy pomocy obiektu StandardScaler. Różnicę przed i po standaryzacji przedstawiłem za pomocą histogramów.

**Kod źródłowy:**

**A screen shot of a computer code

AI-generated content may be incorrect.**

**Wynik:**

**A group of graphs with numbers

AI-generated content may be incorrect.**

1. **Przeprowadzić normalizację wartości atrybutów (przy użyciu różnych norm). Wybraną metodą zwizualizować różnice.**

Normalizację wartości atrybutów wykonałem przy pomocy obiektu Normalizer. Posłużyłem się dwiema normami:

- L1: suma wartości bezwzględnych dla każdego wektora cech równa 1.

- L2: długość każdego wektora cech równa 1.

Różnicę przed i po normalizacji przedstawiłem za pomocą histogramów oraz wyświetliłem w konsoli.

**Kod źródłowy:**

**A screen shot of a computer program

AI-generated content may be incorrect.**

**Wynik:**

**A group of graphs with different colored lines

AI-generated content may be incorrect.**

**A screenshot of a computer screen

AI-generated content may be incorrect.**

1. **Oczyścić dane (uzupełnić/usunąć brakujące wartości, usunąć duplikaty).**

W celu wykonania zadania ponownie posłużyłem się zbiorem danych Iris, lecz tym razem wprowadziłem sztuczne modyfikacje, aby móc następnie oczyścić dane.

Pierwszą modyfikacją było ustawienie 4 pierwszych rekordów atrybutu ‘sepal length (cm)’ na None, aby zasymulować brakujące wartości:

**Kod źródłowy:**

**A screen shot of a computer

AI-generated content may be incorrect.**

**Wynik:**

A screenshot of a computer

AI-generated content may be incorrect.

Następnie brakujące dane postanowiłem dodać jako średnią z wartości danego atrybutu.

**Kod źródłowy:**

**A black screen with colorful text

AI-generated content may be incorrect.**

**Wynik:**

A screenshot of a computer

AI-generated content may be incorrect.

Alternatywnie sprawdziłem również metodę usuwania wektorów cech, gdzie znajdują się niekompletne dane. W tym celu raz jeszcze zmanipulowałem dane wprowadzając brakujące wartości i przystąpiłem do czyszczenia.

**Kod źródłowy:**

**A computer screen shot of code

AI-generated content may be incorrect.**

**Wynik:**

A screenshot of a computer

AI-generated content may be incorrect.

Ostatnią z metod czyszczenia danych było usunięcie duplikatów. W tym celu wczytałem raz jeszcze oryginalny zbiór danych Iris, połączyłem go z samym sobą, aby uzyskać duplikaty. A następnie posłużyłem się funkcją drop\_duplicates() w celu pozbycia się ich.

**Kod źródłowy:**

**A screen shot of a computer program

AI-generated content may be incorrect.**

**Wynik:**

**A screenshot of a computer

AI-generated content may be incorrect.**