

WSI

Ćwiczenie 5

Las losowy

1. Treść zadania

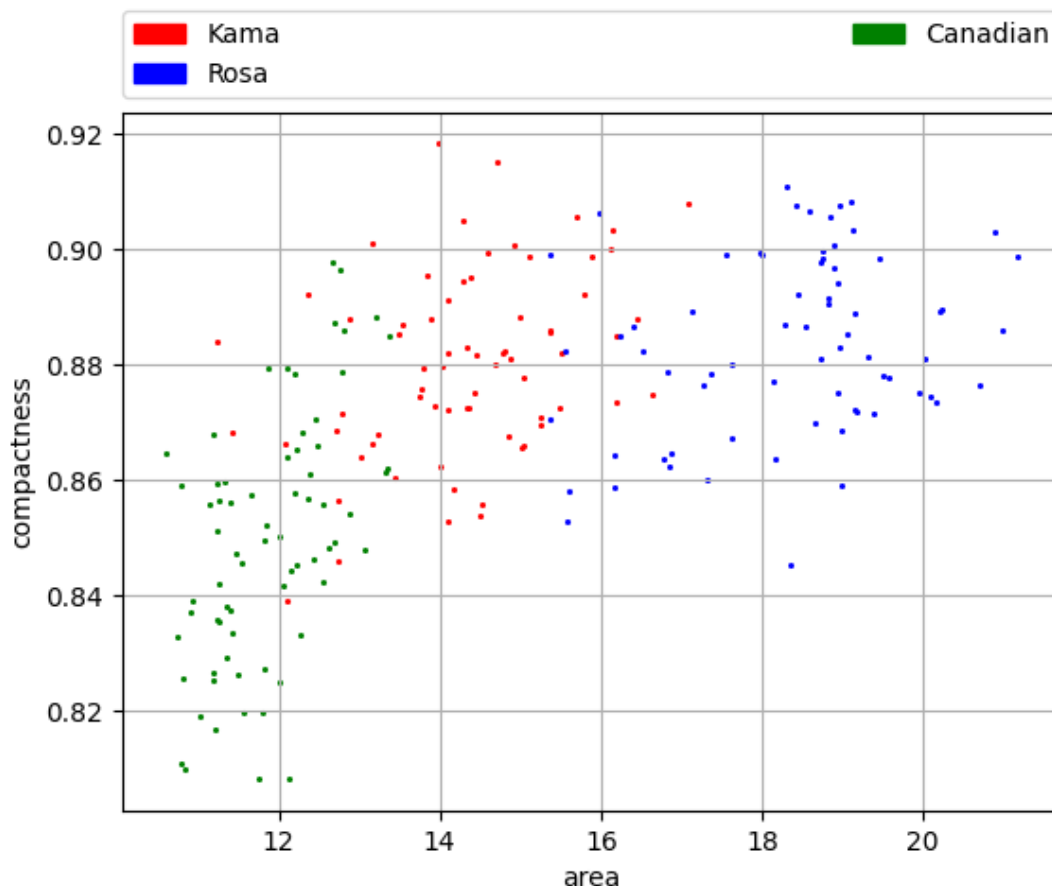
Zadanie polega na implementacji lasu losowego i przeprowadzeniu klasyfikacji metodą k-krotnej walidacji krzyżowej dla zadanego zbioru danych. Należy sprawdzić wygląd zbioru danych oraz wyznaczyć macierz pomyłek z odpowiednimi miarami. Przebadąć wpływ cech wygenerowanych drzew na jakość modelu. Zamieścić przykładowy wygląd drzewa wraz z wartościami węzłów decyzyjnych. Sprawdzić kilka przypadków doboru parametru k dla walidacji krzyżowej.

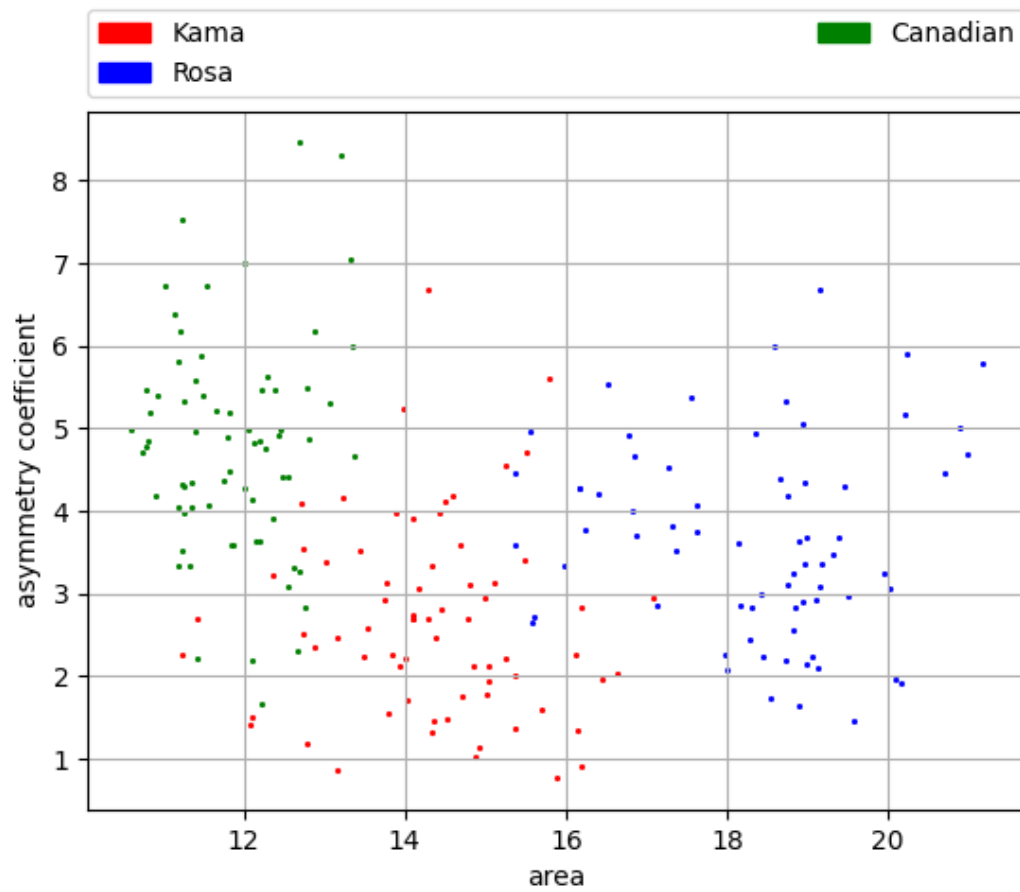
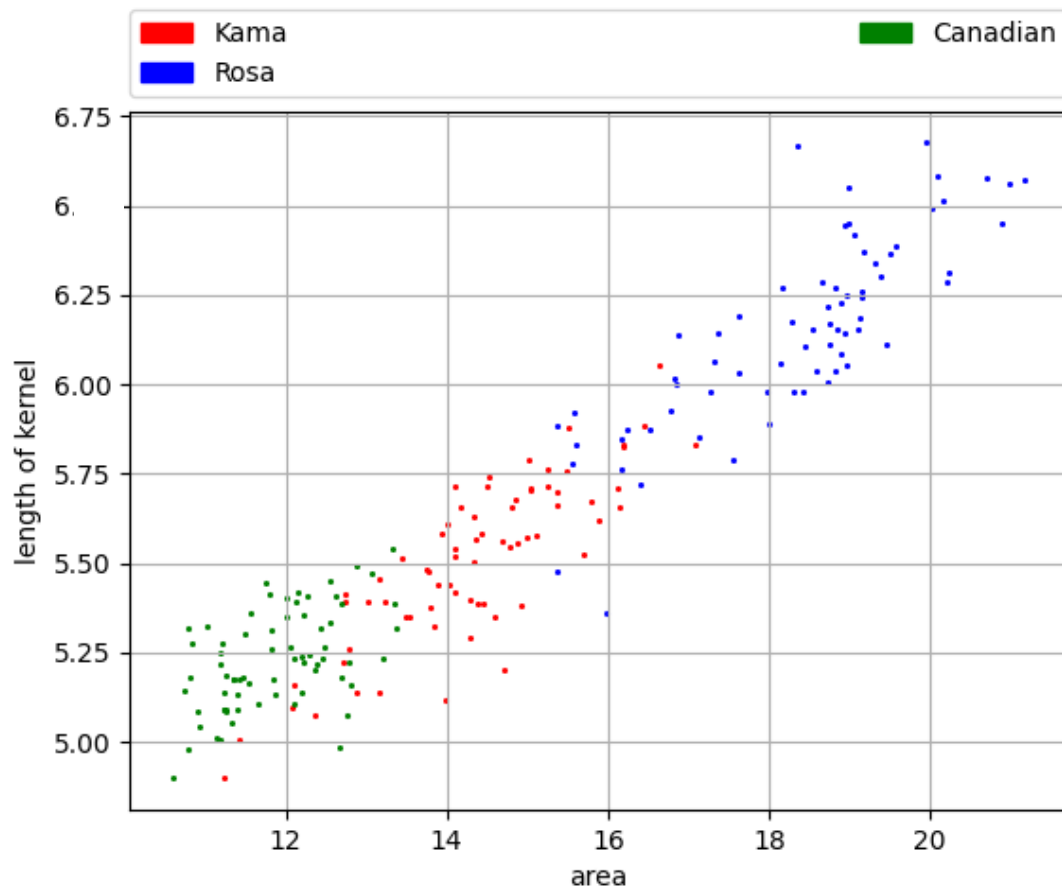
Zbiór danych do klasyfikacji - 3 odmiany pszenicy: Kama, Rosa i Canadian (210 próbek – 3 grupy po 70 elementów; każda próbka ma 7 atrybutów); link do danych:

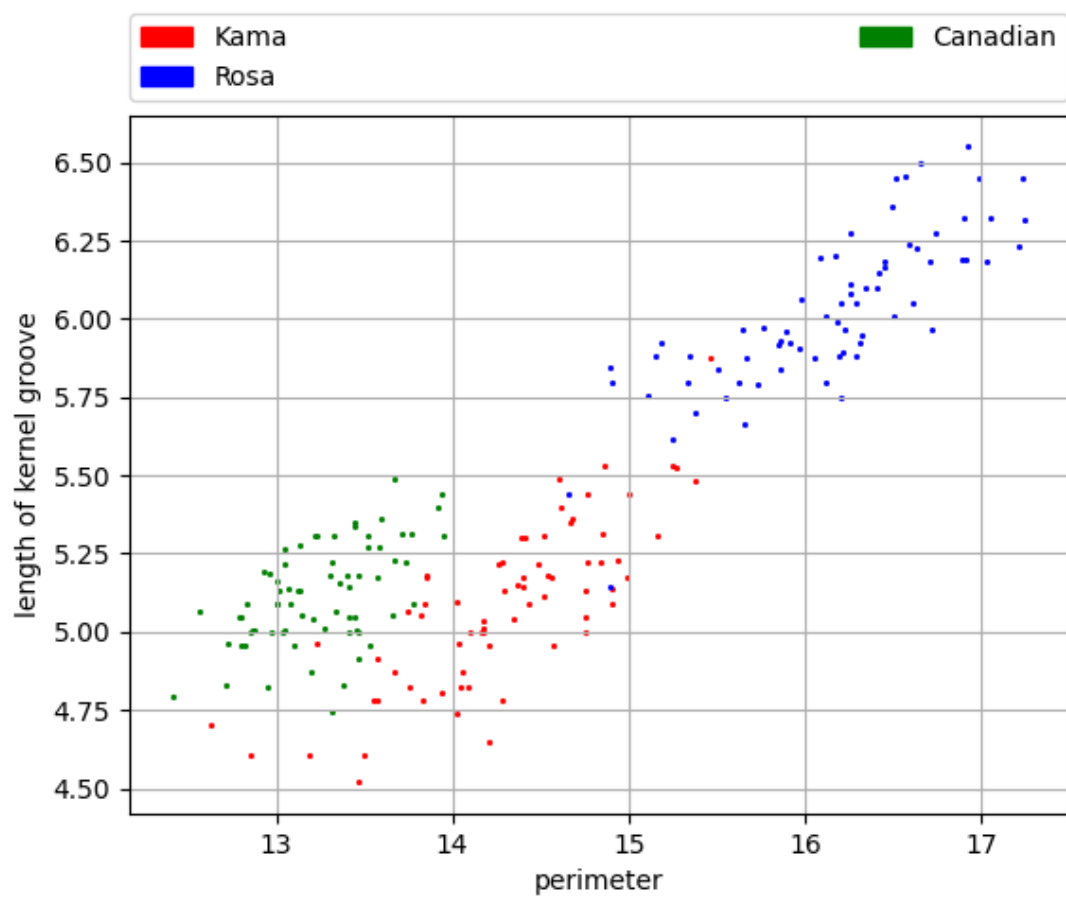
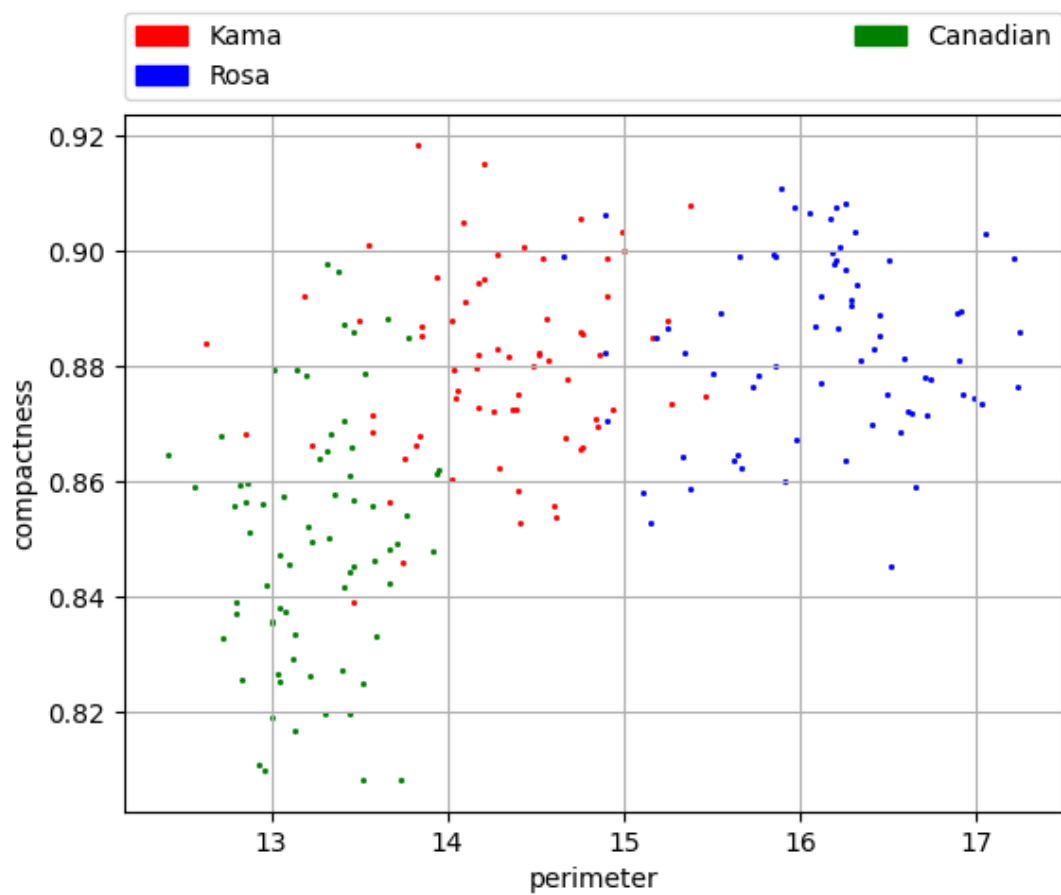
<http://archive.ics.uci.edu/ml/datasets/seeds>.

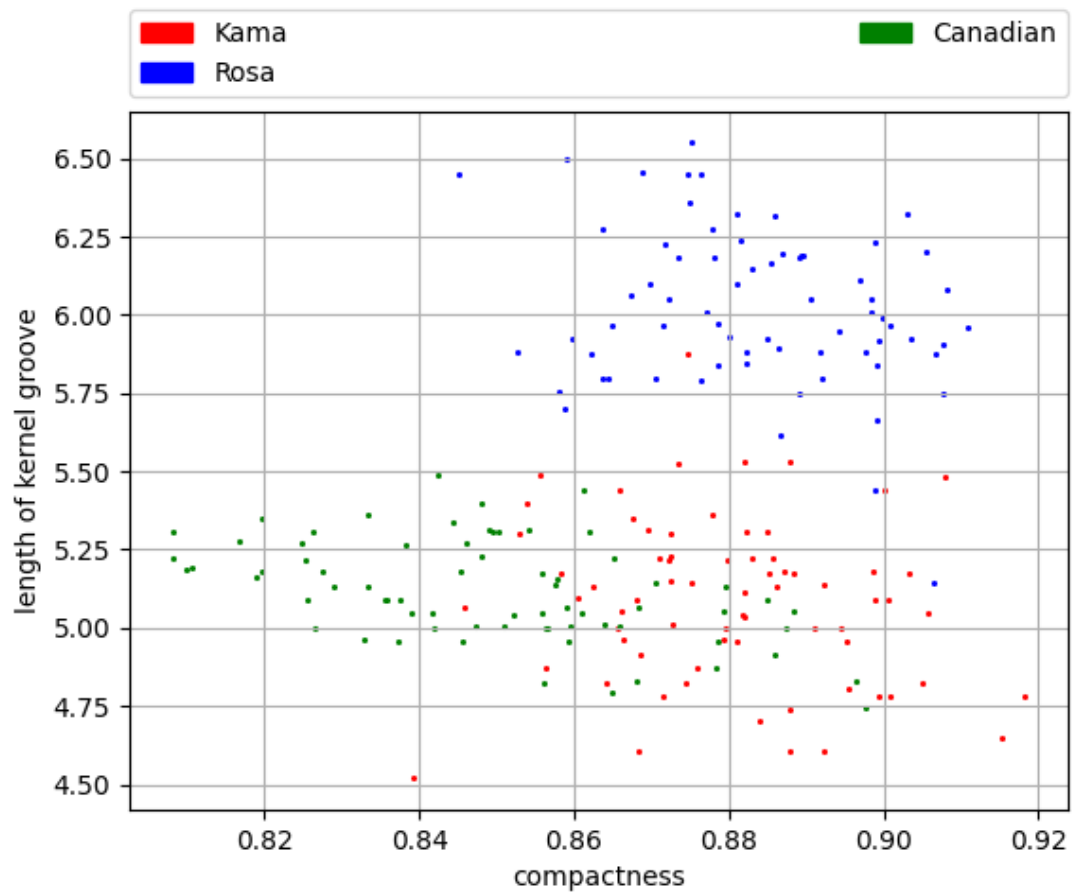
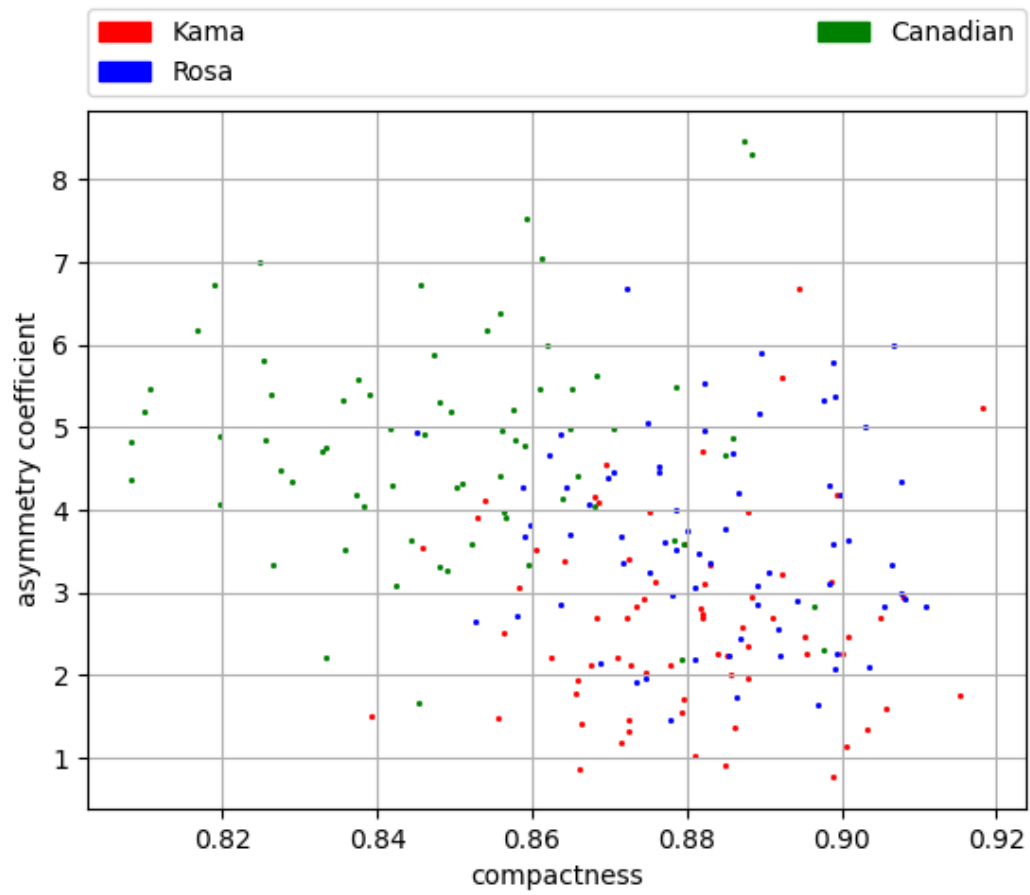
2. Analiza zbioru danych

Separowalność liniowa nie występuje w tym zbiorze danych dla żadnych kombinacji par atrybutów. Sprawdzone dla wszystkich możliwych kombinacji. Poniżej kilka przykładowych wykresów:









3. Opis programu i algorytmu drzew losowych

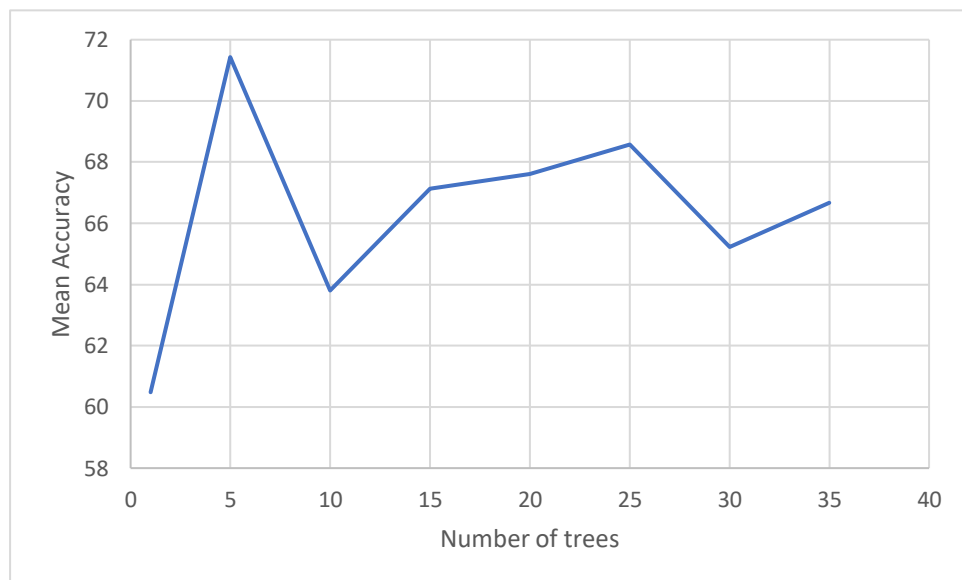
1. Wczytanie i przygotowanie danych
2. Podział zbioru danych za pomocą k-krotnej walidacji krzyżowej; iteracyjny wybór kolejnych zbiorów uczących i testowych; w pętli:
 - a) Uruchomienie algorytmu i otrzymanie wyników klasyfikacji
 - b) Ocena wyników klasyfikacji (porównanie wyników przewidywań z prawdziwymi klasami danych próbek)
 - c) Wyświetlenie wyników (macierz pomyłek, miary)

Algorytm drzew losowych

1. Stworzenie losowego (z powtórzeniami) zbioru danych – „bootstrapped dataset”
2. Wybór atrybutu najlepiej separującego wybrane próbki na podstawie indeksu Ginniego
3. Wykreślenie go z listy dostępnych atrybutów oraz wybór atrybutu kolejnego
4. Kroki 2 i 3 powtarzamy do osiągnięcia maksymalnej głębokości zadanej parametrem – w ten sposób powstaje drzewo decyzyjne
5. Powyższe kroki powtarzamy wielokrotnie – otrzymujemy las
6. Następnie przepuszczamy każdą próbkę zbioru testowego przez każde drzewo w lesie
7. O klasyfikacji danej próbki do danej klasy decyduje wartość klasy, która najczęściej pojawiła się w lesie

4. Wpływ cech wygenerowanych drzew i parametrów algorytmu na jakość modelu

a) Liczba drzew



Przeprowadziłem kilkakrotnie obserwacje wpływu ilości drzew w modelu na jego jakość. Z moich obserwacji wynika, że model osiągał najlepsze rezultaty (ok. 71% dokładności przewidywań) dla liczby drzew równej 5. Dla większych ilości drzew jakość przewidywań ulegała pogorszeniu (64-68%). Z kolei dla liczby drzew mniejszej niż 5, jakość wzrastała wraz z ilością, aż do maksimum w punkcie 5 (drzew).

Przykładowy wynik obserwacji:

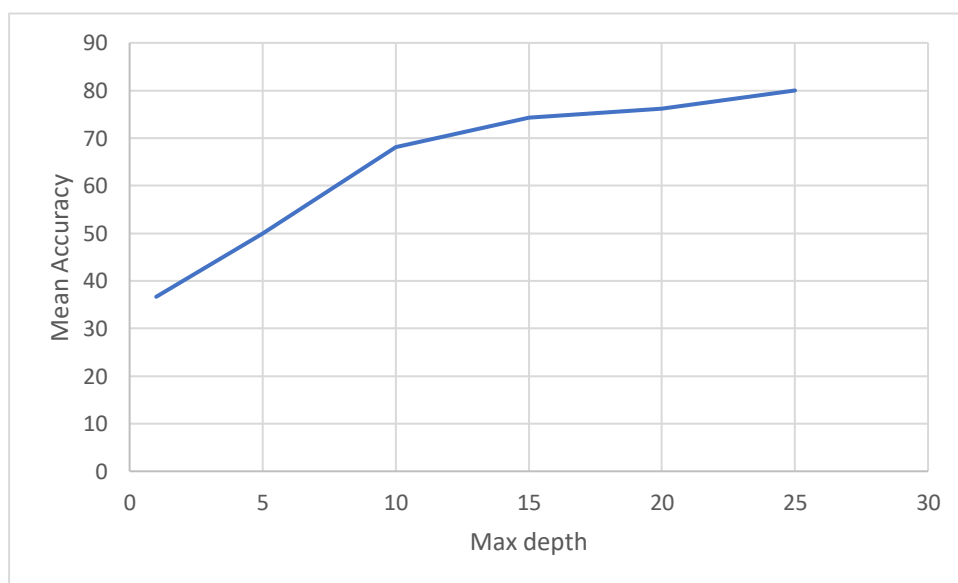
```
Predicted: [3, 1, 3, 2, 3, 3, 3, 3, 2, 2, 2, 3, 1, 3, 1, 3, 1, 2, 2, 3, 2, 3, 2, 3, 2, 1, 1, 2, 2, 2, 2, 1, 2, 3, 3, 3, 2, 2, 3, 2, 1, 2]
Actual:    [3, 1, 3, 2, 3, 3, 3, 3, 2, 1, 2, 3, 1, 3, 1, 1, 1, 2, 2, 1, 2, 3, 1, 3, 2, 1, 1, 2, 2, 1, 2, 1, 1, 1, 3, 3, 2, 1, 1, 1, 1, 2]
Predicted: [2, 2, 1, 2, 2, 2, 3, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 1, 2, 2, 1, 2, 2, 1, 2, 1, 2, 1, 2, 2, 3, 2, 2, 1, 3, 2, 1, 2, 2, 3, 3, 2]
Actual:    [2, 1, 1, 2, 2, 2, 3, 3, 2, 1, 2, 2, 2, 1, 2, 3, 1, 1, 2, 1, 2, 2, 1, 2, 2, 1, 2, 1, 2, 3, 2, 3, 2, 1, 2, 1, 3, 3, 1, 3, 2]
Predicted: [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 3, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
Actual:    [2, 2, 3, 1, 2, 1, 3, 1, 3, 1, 1, 3, 1, 3, 1, 3, 2, 2, 2, 3, 1, 3, 2, 2, 2, 3, 1, 3, 3, 2, 2, 1, 2, 1, 1, 2, 3, 3, 2, 1, 2, 2, 3, 1, 2]
Predicted: [2, 2, 2, 2, 2, 2, 3, 2, 1, 2, 2, 3, 3, 2, 3, 2, 3, 3, 1, 2, 2, 3, 2, 2, 1, 2, 2, 3, 3, 2, 1, 1, 2, 2, 2, 2, 2, 1, 3, 2, 2, 2]
Actual:    [2, 2, 1, 1, 1, 2, 3, 1, 1, 1, 3, 3, 1, 2, 3, 1, 2, 3, 2, 3, 3, 1, 1, 1, 3, 2, 2, 3, 1, 1, 3, 3, 1, 3, 3, 2, 1, 1, 3, 2, 1, 1, 3, 2, 1]
Predicted: [3, 2, 2, 3, 3, 2, 2, 3, 2, 2, 3, 1, 3, 2, 2, 2, 3, 1, 3, 2, 2, 3, 2, 2, 2, 3, 3, 2, 2, 2, 2, 3, 1, 2, 3, 2, 2, 3, 2, 3, 2, 2, 2]
Actual:    [3, 3, 2, 3, 3, 3, 3, 3, 2, 2, 3, 2, 3, 1, 1, 2, 3, 1, 3, 3, 2, 1, 1, 3, 2, 3, 3, 1, 3, 3, 2, 3, 1, 1, 2, 3, 2, 3, 2, 3, 2, 3, 1, 2, 1]
Trees: 1
Scores: [76.19047619047619, 78.57142857142857, 30.952380952380953, 52.38095238095239, 64.28571428571429]
Mean Accuracy: 60.47619047619048
Predicted: [3, 1, 2, 3, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 3, 3, 2, 3, 2, 1, 2, 2, 2, 2, 2, 2, 2, 1, 3, 1, 2, 2, 2, 2, 2, 2, 2, 2]
Actual:    [3, 1, 2, 3, 2, 1, 2, 2, 3, 2, 2, 1, 2, 2, 1, 2, 1, 3, 3, 1, 3, 2, 1, 1, 1, 1, 1, 2, 1, 2, 2, 2, 2, 3, 1, 1, 2, 2, 1, 3, 2, 2, 1]
Predicted: [3, 2, 2, 3, 3, 3, 3, 2, 3, 2, 2, 3, 1, 1, 3, 1, 3, 1, 2, 3, 2, 3, 3, 3, 2, 3, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 3, 2, 2, 3, 3]
Actual:    [3, 2, 2, 3, 3, 1, 3, 1, 3, 1, 2, 3, 1, 1, 3, 1, 3, 1, 1, 3, 1, 3, 3, 3, 2, 3, 1, 1, 1, 1, 3, 2, 3, 1, 1, 3, 2, 1, 1, 3, 2, 1, 3, 2, 1]
Predicted: [2, 1, 3, 3, 2, 2, 3, 3, 2, 2, 2, 3, 1, 2, 2, 3, 2, 2, 1, 2, 2, 1, 3, 2, 2, 3, 3, 3, 2, 2, 1, 2, 3, 3, 2, 3, 2, 3, 2, 3, 2, 2, 2]
Actual:    [2, 1, 3, 3, 2, 2, 3, 3, 2, 2, 2, 1, 3, 1, 1, 2, 3, 1, 2, 1, 1, 2, 1, 3, 2, 2, 3, 3, 3, 2, 1, 1, 1, 3, 1, 3, 3, 3, 3, 3, 2, 1, 1]
```

```

Predicted: [2, 2, 2, 2, 1, 2, 2, 3, 2, 2, 3, 2, 2, 2, 2, 1, 2, 3, 2, 3, 3, 2, 2, 2, 2, 1, 3, 2, 2, 3, 3, 2, 2, 1, 2, 2, 3, 3, 1, 2, 2, 2]
Actual:    [2, 3, 2, 2, 1, 3, 2, 3, 1, 2, 3, 3, 2, 1, 2, 1, 2, 3, 1, 3, 3, 1, 1, 3, 2, 1, 1, 3, 1, 1, 3, 3, 3, 1, 1, 1, 2, 3, 3, 1, 1, 2]
Predicted: [2, 2, 2, 2, 3, 2, 2, 3, 2, 2, 3, 3, 3, 3, 1, 3, 2, 2, 3, 1, 3, 2, 1, 2, 2, 2, 2, 2, 3, 2, 2, 2, 2, 2, 3, 3, 3, 2, 2, 2, 2, 2]
Actual:    [1, 2, 1, 2, 3, 2, 1, 3, 1, 3, 1, 3, 3, 1, 3, 2, 1, 3, 1, 3, 1, 1, 2, 2, 2, 1, 1, 2, 2, 3, 2, 2, 3, 3, 3, 1, 1, 3, 3, 1, 1, 3, 1]
Predicted: [2, 3, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2]
Actual:    [2, 3, 2, 3, 1, 2, 2, 1, 1, 3, 3, 2, 2, 2, 1, 2, 1, 3, 3, 1, 1, 3, 1, 1, 1, 2, 1, 2, 3, 1, 3, 2, 1, 3, 2, 3, 2, 2, 3, 2, 2, 3, 1, 2, 2]
Predicted: [2, 2, 2, 2, 3, 2, 2, 1, 3, 2, 2, 3, 3, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2]
Actual:    [1, 2, 2, 2, 3, 2, 2, 2, 1, 3, 1, 2, 3, 3, 2, 2, 3, 2, 2, 2, 2, 3, 1, 2, 2, 1, 2, 3, 1, 1, 3, 2, 2, 3, 3, 1, 1, 3, 3, 3, 3, 3, 1]
Predicted: [2, 2, 3, 3, 2, 3, 2, 2, 3, 2, 2, 3, 2, 3, 2, 2, 1, 3, 1, 1, 2, 2, 3, 2, 2, 3, 2, 2, 3, 2, 2, 2, 1, 2, 3, 1, 1, 2, 2, 1, 2, 3, 3, 1, 3, 3]
Actual:    [2, 1, 3, 3, 2, 3, 2, 2, 3, 3, 1, 3, 1, 3, 2, 2, 1, 3, 1, 3, 1, 2, 3, 2, 1, 3, 1, 1, 2, 3, 2, 1, 3, 1, 1, 2, 3, 1, 1, 1, 1, 2, 3, 3, 1, 3, 3]
Trees: 15
Scores: [59.523809523809526, 64.28571428571429, 52.38095238095239, 80.95238095238095, 78.57142857142857]
Mean Accuracy: 67.14285714285714
Predicted: [2, 2, 2, 2, 2, 2, 2, 3, 3, 2, 2, 2, 3, 2, 1, 2, 2, 2, 2, 2, 2, 2, 3, 2, 1, 2, 2, 3, 2, 2, 2, 2, 2, 1, 2, 2, 2, 1, 2, 2, 2, 3, 2, 2]
Actual:    [1, 2, 2, 2, 3, 2, 1, 3, 3, 1, 3, 1, 3, 2, 3, 2, 2, 1, 2, 2, 1, 3, 1, 3, 1, 2, 3, 2, 2, 1, 2, 3, 1, 2, 3, 1, 2, 1, 1, 2, 2, 3, 3, 2, 1]
Predicted: [2, 1, 3, 2, 3, 2, 3, 3, 2, 2, 1, 3, 3, 3, 2, 3, 3, 2, 2, 3, 2, 1, 2, 1, 1, 3, 3, 1, 2, 3, 2, 2, 2, 3, 2, 3, 2, 2, 2, 3, 2, 2, 2]
Actual:    [2, 1, 3, 3, 3, 2, 3, 1, 1, 2, 1, 1, 3, 3, 2, 3, 3, 2, 3, 3, 2, 3, 2, 1, 2, 1, 1, 3, 3, 1, 1, 3, 2, 2, 1, 3, 3, 3, 3, 2, 2, 3, 1, 1]
Predicted: [2, 3, 2, 2, 2, 2, 3, 2, 2, 2, 3, 3, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 2, 1, 3, 2, 2, 2, 1, 3, 2, 3, 3]
Actual:    [2, 1, 1, 3, 2, 2, 3, 2, 2, 1, 2, 3, 3, 3, 1, 3, 2, 3, 1, 3, 1, 1, 3, 1, 3, 3, 3, 3, 2, 1, 1, 1, 3, 2, 2, 2, 1, 3, 2, 3, 3, 3]
Predicted: [2, 2, 2, 2, 1, 2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 3, 2, 2, 2, 3, 1, 3, 2, 3, 2, 3, 2, 3, 3, 2, 3, 3, 2, 2, 3, 1, 1, 2, 1, 1, 1, 2, 1]
Actual:    [1, 2, 2, 2, 1, 1, 1, 1, 1, 2, 2, 1, 2, 3, 1, 1, 3, 2, 3, 1, 1, 3, 3, 1, 3, 3, 1, 3, 3, 1, 3, 2, 3, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2]
Predicted: [2, 2, 2, 3, 3, 3, 2, 2, 2, 1, 2, 3, 2, 2, 2, 3, 2, 2, 2, 2, 3, 3, 2, 2, 2, 3, 3, 2, 2, 2, 3, 3, 2, 2, 2, 2, 3, 3, 2, 2, 3, 3, 2]
Actual:    [2, 2, 1, 3, 3, 3, 2, 1, 3, 1, 2, 2, 3, 2, 2, 1, 2, 3, 1, 2, 2, 1, 1, 3, 3, 2, 2, 1, 2, 3, 3, 1, 1, 1, 3, 3, 2, 2, 1, 2, 3, 3, 2]
Trees: 20
Scores: [59.523809523809526, 76.19047619047619, 71.42857142857143, 59.523809523809526, 71.42857142857143]
Mean Accuracy: 67.61904761904762
Predicted: [2, 2, 3, 3, 2, 3, 3, 3, 2, 2, 2, 2, 2, 3, 3, 2, 1, 2, 2, 2, 3, 2, 2, 2, 2, 3, 2, 3, 2, 2, 2, 1, 2, 1, 3, 2, 3, 2, 2, 2, 2, 3, 2]
Actual:    [2, 2, 3, 3, 2, 1, 3, 3, 3, 2, 1, 2, 3, 3, 1, 1, 1, 2, 2, 3, 2, 2, 2, 1, 3, 2, 3, 1, 2, 1, 2, 1, 3, 2, 3, 1, 3, 2, 3, 1, 3, 2, 3, 2]
Predicted: [2, 2, 2, 3, 3, 2, 2, 2, 2, 2, 3, 2, 2, 3, 2, 3, 2, 2, 2, 3, 2, 3, 2, 2, 2, 3, 2, 3, 2, 3, 2, 1, 2, 2, 3, 2, 2, 3, 2, 2, 3, 2, 2]
Actual:    [1, 2, 1, 3, 3, 1, 2, 1, 2, 3, 2, 1, 3, 2, 2, 2, 3, 3, 2, 1, 3, 1, 2, 3, 1, 3, 3, 3, 1, 1, 2, 2, 3, 3, 3, 1, 3, 2, 2, 3, 2, 2, 2]
Predicted: [3, 3, 2, 2, 3, 2, 3, 2, 1, 2, 2, 2, 2, 3, 2, 3, 2, 2, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 3, 2, 2, 2, 3, 3, 3, 2, 2, 3, 2, 3, 3, 3]
Actual:    [3, 3, 3, 1, 1, 2, 3, 2, 1, 1, 1, 1, 1, 3, 2, 3, 1, 1, 3, 1, 3, 3, 2, 3, 1, 2, 2, 1, 3, 3, 3, 2, 2, 1, 3, 2, 1, 1, 1, 3, 3, 3, 3]
Predicted: [2, 2, 1, 2, 2, 3, 2, 3, 2, 3, 2, 1, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 1, 2, 2, 2, 1, 1, 2, 2, 2, 1, 2, 2, 3, 2, 3, 2, 2, 3, 1, 2, 2]
Actual:    [2, 1, 1, 2, 2, 3, 3, 3, 2, 1, 2, 2, 1, 2, 2, 1, 1, 1, 2, 1, 1, 2, 3, 1, 1, 1, 2, 1, 1, 2, 1, 1, 2, 1, 3, 2, 3, 2, 2, 3, 1, 2, 3]
Predicted: [3, 1, 1, 1, 2, 2, 1, 2, 3, 2, 3, 2, 3, 2, 2, 2, 2, 3, 2, 3, 1, 2, 2, 3, 2, 3, 2, 3, 2, 2, 1, 1, 2, 1, 3, 3, 3, 3, 3, 3, 2, 3, 2]
Actual:    [3, 2, 1, 1, 1, 2, 1, 1, 2, 3, 3, 3, 3, 2, 1, 1, 2, 3, 1, 3, 1, 2, 3, 1, 2, 3, 1, 2, 3, 1, 2, 3, 1, 2, 1, 1, 2, 3, 3, 3, 3, 3, 2, 3, 2]
Trees: 25
Scores: [78.57142857142857, 61.904761904761905, 59.523809523809526, 71.42857142857143, 71.42857142857143]
Mean Accuracy: 68.57142857142858
Predicted: [1, 3, 2, 2, 2, 2, 2, 2, 1, 1, 2, 3, 2, 2, 2, 2, 1, 3, 2, 2, 2, 2, 3, 2, 3, 3, 2, 3, 3, 2, 2, 2, 1, 2, 3, 1, 1, 3, 2, 3, 2, 3, 2, 2]
Actual:    [1, 3, 1, 3, 2, 2, 2, 1, 3, 3, 3, 2, 2, 2, 1, 3, 3, 2, 2, 3, 3, 2, 3, 3, 2, 3, 3, 2, 3, 2, 1, 1, 1, 2, 3, 1, 1, 2, 1, 3, 2, 1, 2, 3]
Predicted: [2, 1, 2, 2, 2, 3, 2, 2, 2, 3, 1, 2, 2, 3, 2, 2, 3, 3, 2, 3, 2, 3, 3, 2, 3, 3, 2, 1, 2, 2, 2, 2, 2, 3, 2, 3, 3, 3, 2, 3, 3, 3, 2, 2]
Actual:    [3, 3, 1, 1, 1, 2, 3, 2, 2, 3, 3, 1, 1, 1, 3, 2, 2, 3, 3, 1, 1, 3, 2, 3, 3, 3, 1, 2, 3, 2, 2, 2, 3, 3, 3, 2, 2, 3, 3, 3, 1, 3, 3, 1, 1, 2]
Predicted: [2, 2, 2, 2, 3, 3, 3, 2, 2, 2, 2, 3, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 3, 2, 2, 3, 1, 2, 3, 2, 2, 2, 2, 2, 2, 3, 3, 1, 2, 3, 3, 1, 2, 3, 2, 2]
Actual:    [1, 2, 2, 2, 1, 3, 3, 1, 1, 2, 1, 3, 3, 1, 1, 2, 1, 3, 3, 1, 1, 3, 1, 1, 3, 1, 2, 3, 1, 1, 2, 3, 1, 1, 2, 3, 1, 1, 2, 3, 3, 1, 1, 3, 2, 1]
Predicted: [2, 3, 2, 1, 2, 3, 1, 1, 1, 3, 1, 2, 2, 2, 2, 2, 2, 2, 3, 2, 2, 2, 2, 2, 2, 1, 2, 1, 2, 1, 2, 2, 2, 3, 1, 3, 1, 2, 3, 2, 2, 3, 2, 2]
Actual:    [2, 3, 1, 1, 2, 3, 1, 1, 3, 3, 1, 2, 2, 1, 3, 1, 1, 1, 3, 2, 2, 2, 2, 1, 2, 1, 3, 1, 2, 2, 2, 3, 1, 3, 1, 2, 3, 2, 2, 3, 2, 2, 1]
Predicted: [3, 2, 2, 2, 3, 2, 2, 2, 2, 2, 3, 2, 2, 1, 2, 3, 2, 2, 3, 2, 2, 2, 3, 2, 2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 3, 2]
Actual:    [3, 2, 1, 2, 3, 1, 3, 3, 1, 2, 3, 2, 1, 2, 1, 3, 2, 2, 3, 2, 3, 3, 3, 2, 1, 1, 1, 1, 1, 1, 3, 1, 1, 2, 1, 2, 1, 2, 1, 2, 2, 2, 3, 2]
Trees: 30
Scores: [71.42857142857143, 69.04761904761905, 52.38095238095239, 78.57142857142857, 54.761904761904766]
Mean Accuracy: 65.23809523809524
Predicted: [3, 3, 2, 2, 2, 2, 2, 2, 2, 1, 2, 1, 2, 3, 2, 2, 1, 3, 2, 2, 2, 2, 2, 1, 2, 3, 3, 2, 3, 2, 2, 2, 1, 2, 2, 3, 2, 1, 3, 2, 2, 2]
Actual:    [3, 3, 2, 1, 3, 2, 2, 2, 2, 1, 3, 1, 1, 3, 1, 3, 2, 1, 1, 3, 2, 3, 1, 3, 1, 2, 3, 3, 2, 3, 1, 3, 2, 1, 2, 2, 3, 2, 1, 3, 2, 3]
Predicted: [3, 3, 3, 3, 3, 2, 2, 2, 3, 2, 1, 2, 3, 2, 3, 3, 2, 3, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 2, 2, 2, 3, 2, 2, 1, 3, 2, 3, 1, 2, 2, 2, 2]
Actual:    [3, 3, 3, 3, 3, 3, 2, 3, 1, 2, 2, 3, 1, 3, 1, 2, 3, 2, 2, 2, 1, 1, 2, 2, 2, 1, 3, 2, 2, 2, 3, 1, 1, 1, 3, 1, 3, 1, 1, 1, 2, 1]
Predicted: [2, 2, 2, 3, 2, 3, 2, 2, 2, 2, 2, 2, 3, 2, 2, 2, 3, 2, 2, 2, 3, 3, 2, 2, 2, 2, 3, 3, 2, 2, 2, 2, 2, 2, 3, 2, 2, 2, 1, 2, 2, 1, 2]
Actual:    [1, 2, 2, 2, 3, 3, 3, 2, 2, 1, 2, 3, 2, 3, 3, 3, 1, 1, 2, 2, 3, 3, 3, 2, 3, 1, 2, 2, 3, 3, 1, 2, 2, 1, 1, 2, 2, 1, 1, 1, 1, 1, 2]
Predicted: [1, 2, 3, 2, 3, 2, 1, 2, 2, 2, 2, 2, 2, 3, 2, 2, 3, 1, 2, 2, 2, 2, 2, 3, 2, 3, 2, 3, 2, 2, 1, 2, 2, 3, 3, 3, 2, 2, 2, 2, 2, 3]
Actual:    [1, 2, 3, 1, 3, 2, 1, 3, 2, 2, 2, 1, 1, 3, 3, 1, 1, 1, 1, 1, 2, 1, 1, 2, 3, 2, 3, 3, 2, 2, 1, 2, 2, 3, 3, 3, 2, 1, 3, 1, 3, 1, 3]
Predicted: [3, 2, 2, 3, 2, 2, 2, 2, 2, 3, 3, 3, 2, 3, 1, 3, 2, 3, 2, 2, 2, 3, 3, 2, 2, 1, 3, 2, 3, 2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2]
Actual:    [3, 1, 1, 3, 2, 2, 2, 1, 1, 1, 3, 3, 3, 3, 1, 3, 2, 3, 1, 1, 1, 1, 3, 1, 3, 1, 3, 1, 3, 3, 3, 2, 1, 1, 2, 2, 2, 2, 1, 1, 2, 1, 2, 2]
Trees: 35
Scores: [69.04761904761905, 64.28571428571429, 71.42857142857143, 66.66666666666666, 61.904761904761905]
Mean Accuracy: 66.66666666666667

```

b) Maksymalna głębokość drzew



Przeprowadziłem kilkakrotne obserwacje wpływu głębokości drzew na jakość przewidywań modelu. Z moich obserwacji wynika, że im głębsza głębokość drzew w lesie, tym dokładniejszy wynik przewidywań. Jest to zależność liniowa.

Przykładowy wynik obserwacji:

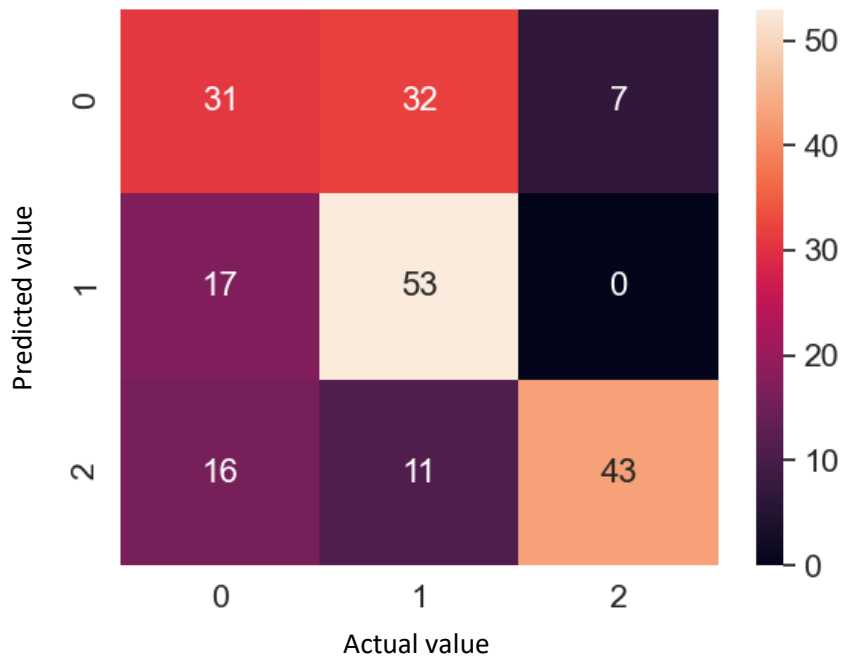
[illegible]

ltd.

5. Przykładowy wynik działania algorytmu wraz z macierzą pomyłek oraz różnymi miarami

*Dla zmiennej ilości drzew

Liczba drzew = 1



```
Predicted: [3, 1, 3, 2, 3, 3, 3, 3, 2, 2, 2, 3, 1, 3, 1, 3, 1, 2, 2, 3, 2, 3, 2, 3, 2, 1, 1, 2, 2, 2, 1, 2, 3, 3, 3, 2, 2, 3, 2, 1, 2]
Actual:     [3, 1, 3, 2, 3, 3, 3, 3, 2, 1, 2, 3, 1, 3, 1, 1, 1, 2, 2, 1, 2, 3, 1, 3, 2, 1, 1, 2, 2, 1, 2, 1, 1, 1, 3, 3, 2, 1, 1, 1, 1, 2]
Predicted: [2, 2, 1, 2, 2, 2, 3, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 1, 2, 2, 1, 2, 2, 1, 2, 1, 2, 2, 2, 3, 2, 2, 1, 3, 2, 1, 2, 2, 3, 2, 3, 2]
Actual:     [2, 1, 1, 2, 2, 2, 3, 3, 2, 1, 2, 2, 2, 1, 2, 3, 1, 1, 2, 1, 2, 2, 1, 2, 1, 2, 3, 2, 3, 2, 2, 3, 3, 2, 1, 2, 1, 3, 3, 1, 3, 2]
Predicted: [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 3, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
Actual:     [2, 2, 3, 1, 2, 1, 3, 1, 3, 1, 1, 3, 1, 3, 2, 2, 2, 2, 3, 1, 3, 2, 1, 3, 3, 3, 2, 2, 1, 2, 1, 1, 2, 3, 3, 2, 1, 2, 2, 3, 1, 2]
Predicted: [2, 2, 2, 2, 2, 2, 3, 2, 1, 2, 2, 3, 3, 2, 3, 2, 3, 3, 1, 2, 2, 3, 2, 2, 1, 2, 2, 3, 3, 2, 1, 1, 2, 2, 2, 2, 2, 1, 3, 2, 2, 2]
Actual:     [2, 2, 1, 1, 1, 2, 3, 1, 1, 1, 3, 3, 1, 2, 3, 2, 3, 3, 1, 1, 1, 3, 2, 2, 3, 1, 1, 3, 3, 1, 3, 3, 2, 1, 1, 3, 2, 1, 1, 3, 2, 1]
Predicted: [3, 2, 2, 3, 3, 2, 2, 3, 2, 2, 3, 1, 3, 2, 2, 2, 3, 1, 3, 2, 2, 3, 2, 2, 2, 2, 3, 3, 2, 2, 2, 2, 3, 1, 2, 3, 2, 3, 2, 3, 2, 2]
Actual:     [3, 3, 2, 3, 3, 3, 3, 3, 3, 2, 2, 3, 2, 3, 1, 1, 2, 3, 1, 3, 3, 2, 2, 1, 1, 3, 2, 3, 3, 1, 3, 3, 2, 3, 1, 2, 3, 2, 3, 2, 3, 1, 2, 1]

Confusion matrix measures
Accuracy: 0.6047619047619047
Micro Precision: 0.6047619047619047
Micro Recall: 0.6047619047619047
Micro F1-score: 0.6047619047619047

Macro Precision: 0.6321527777777778
Macro Recall: 0.6047619047619047
Macro F1-score: 0.6059691502327719

Weighted Precision: 0.6321527777777778
Weighted Recall: 0.6047619047619047
Weighted F1-score: 0.6059691502327719
```

Classification Report

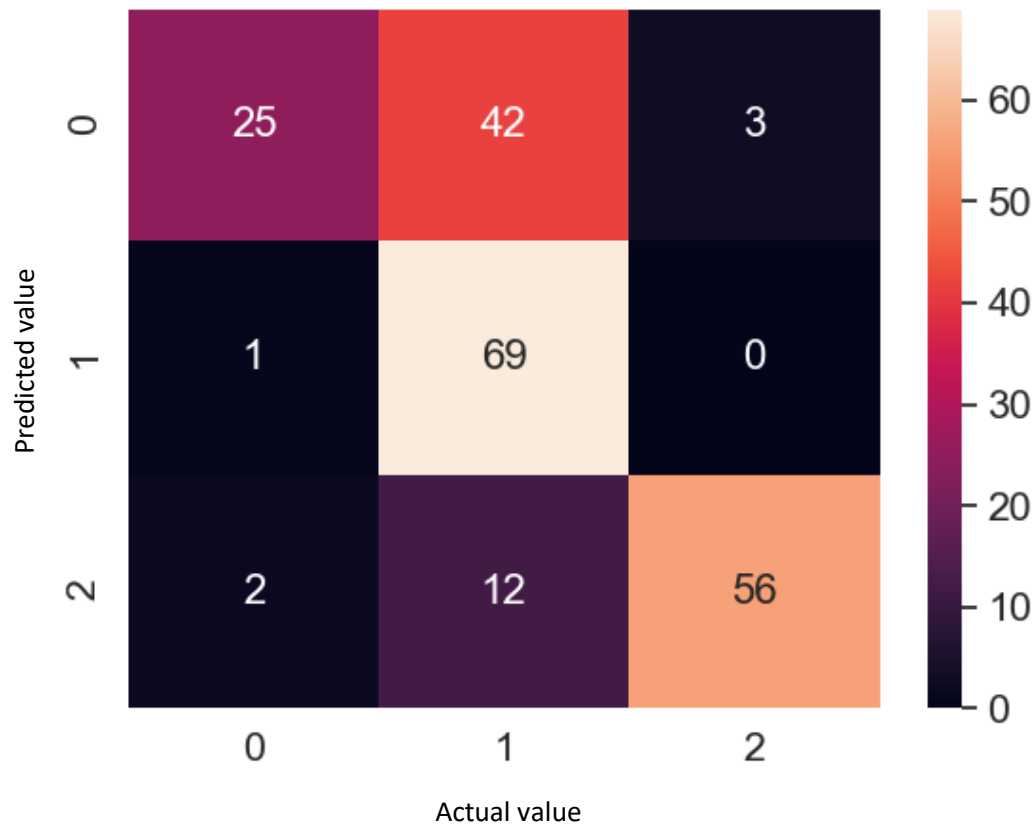
	precision	recall	f1-score	support
1	0.48	0.44	0.46	70
2	0.55	0.76	0.64	70
3	0.86	0.61	0.72	70
accuracy			0.60	210
macro avg	0.63	0.60	0.61	210
weighted avg	0.63	0.60	0.61	210

Trees: 1

Scores: [76.19047619047619, 78.57142857142857, 30.952380952380953, 52.38095238095239, 64.28571428571429]

Mean Accuracy: 60.47619047619048

Liczba drzew = 5



```
Predicted: [3, 1, 2, 3, 2, 1, 2, 2, 2, 2, 2, 2, 1, 2, 2, 3, 3, 2, 3, 2, 2, 1, 2, 2, 2, 2, 2, 2, 1, 3, 1, 2, 2, 2, 2, 2, 2, 2, 2]
Actual:     [3, 1, 2, 3, 2, 1, 2, 2, 3, 2, 2, 1, 2, 2, 1, 2, 1, 3, 3, 1, 3, 2, 1, 1, 1, 1, 2, 1, 2, 2, 2, 2, 3, 1, 1, 2, 2, 1, 3, 2, 2, 1]
Predicted: [3, 2, 2, 3, 3, 3, 3, 2, 3, 2, 2, 3, 1, 1, 3, 1, 3, 1, 2, 3, 2, 3, 3, 3, 2, 3, 1, 2, 2, 2, 2, 2, 2, 2, 2, 1, 3, 2, 2, 3, 3, 1]
Actual:     [3, 2, 2, 3, 3, 3, 3, 1, 3, 1, 2, 3, 1, 1, 3, 1, 3, 1, 3, 1, 3, 3, 3, 2, 3, 1, 1, 1, 1, 1, 3, 2, 3, 1, 2, 1, 3, 2, 2, 3, 1]
Predicted: [2, 1, 3, 3, 2, 2, 3, 3, 2, 2, 2, 2, 3, 1, 2, 2, 3, 2, 2, 1, 2, 2, 1, 3, 2, 2, 3, 3, 3, 2, 2, 1, 2, 3, 3, 2, 3, 2, 3, 2, 2, 2]
Actual:     [2, 1, 3, 3, 2, 2, 3, 3, 2, 2, 2, 1, 3, 1, 1, 2, 3, 1, 2, 1, 1, 2, 1, 3, 2, 2, 3, 3, 3, 2, 1, 1, 1, 3, 1, 3, 3, 3, 3, 2, 1, 1]
Predicted: [2, 2, 3, 3, 3, 2, 2, 2, 2, 1, 3, 3, 2, 2, 2, 2, 2, 2, 3, 3, 2, 2, 2, 3, 2, 1, 3, 2, 2, 2, 1, 3, 2, 2, 2, 1, 3, 2, 2, 2, 3, 2, 1]
Actual:     [3, 2, 3, 3, 3, 3, 2, 3, 2, 3, 2, 2, 1, 3, 3, 2, 2, 2, 1, 2, 3, 3, 2, 2, 3, 3, 1, 3, 1, 3, 3, 1, 2, 1, 1, 3, 2, 3, 1, 1, 3, 2, 1]
Predicted: [2, 2, 2, 1, 2, 2, 2, 2, 2, 2, 3, 2, 3, 2, 2, 2, 2, 2, 2, 1, 2, 3, 1, 1, 3, 3, 2, 2, 2, 1, 3, 3, 2, 1, 3, 2, 2, 1, 3, 2, 2, 2]
Actual:     [3, 2, 2, 1, 2, 2, 2, 1, 2, 1, 2, 3, 2, 3, 2, 1, 1, 2, 1, 1, 1, 1, 1, 3, 3, 1, 3, 3, 1, 2, 2, 1, 3, 3, 2, 1, 3, 2, 1, 3, 3, 2, 2, 2]

Confusion matrix measures
Accuracy: 0.7142857142857143
Micro Precision: 0.7142857142857143
Micro Recall: 0.7142857142857143
Micro F1-score: 0.7142857142857143

Macro Precision: 0.8009950983287073
Macro Recall: 0.7142857142857144
Macro F1-score: 0.6978156808773234

Weighted Precision: 0.8009950983287073
Weighted Recall: 0.7142857142857143
Weighted F1-score: 0.6978156808773234
```

```
Classification Report
```

	precision	recall	f1-score	support
1	0.89	0.36	0.51	70
2	0.56	0.99	0.72	70
3	0.95	0.80	0.87	70
accuracy			0.71	210
macro avg	0.80	0.71	0.70	210
weighted avg	0.80	0.71	0.70	210

Trees: 5

Scores: [69.04761904761905, 69.04761904761905, 73.80952380952381, 73.80952380952381, 71.42857142857143]

Mean Accuracy: 71.42857142857143

6. Przykładowy wygląd drzewa decyzyjnego

Wyrysowany ręcznie dla przykładowego drzewa o maksymalnej głębokości 3 na podstawie słownika wartości wygenerowanego przez algorytm:

```
{'index': 6, 'value': 4.738, 'left': 1, 'right': {'index': 1, 'value': 13.57, 'left': {'index': 2, 'value': 0.8167, 'left': 3, 'right': 3}, 'right': {'index': 1, 'value': 13.71, 'left': 1, 'right': 2}}}
```

Index – indeks atrybutu obserwacji, na podstawie którego dokonujemy decyzji

Value – wartość rozdzielająca

Left, right – dzieci (jeśli sama liczba, to znaczy, że jest to liść)

