Metody Inteligencji Obliczeniowej – Projekt

Dostrajanie klasyfikatora Fuzzy Logic z użyciem Colliding Bodies Optimization

Piotr Deda, Aleksander Kluczka, Mariusz Biegański

1. Opis projektu

Celem projektu jest wykorzystanie układu Fuzzy Logic do klasyfikacji zbiorów danych benchmarkowych IRIS, WINE oraz SEEDS, które znajdują się na stronie http://archive.ics.uci.edu/ml/datasets.php.

Należy zbudować układ Fuzzy Logic jakąś ogólną zasadą budowy układu FL, a następnie doprecyzować parametry za pomocą wybranego algorytmu, a na koniec porównać oba układy z użyciem algorytmu Cross-Validation (CV-5).

2. Opis algorytmu optymalizacji

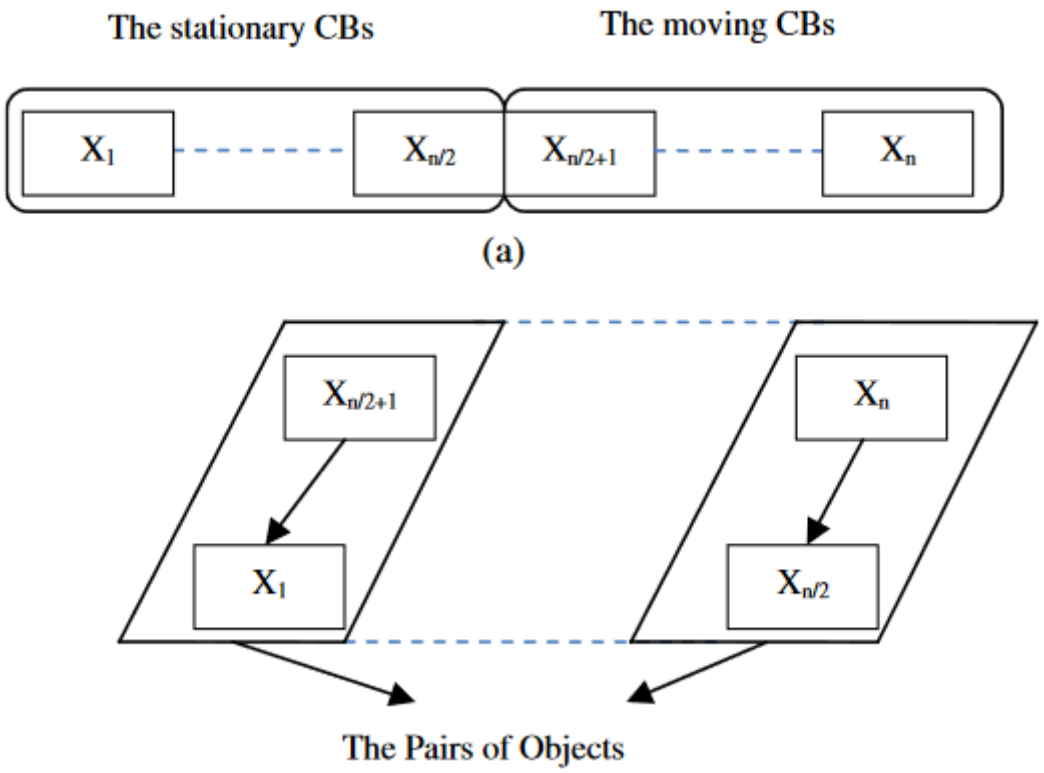
Colliding Bodies Optimization (CBO) to metaheurystyczna metoda optymalizacji oparta na jednowymiarowych kolizjach między ciałami. Ciała reprezentują zestawy wartości, które mają większą masę dla lepszych wyników funkcji dopasowania (Wzór 1).

$$m_k = \frac{\frac{1}{fit(k)}}{\sum_{i=1}^{n} \frac{1}{fit(i)}}, \quad k = 1, 2, ..., n \qquad \textit{(Wzór 1)}$$

W każdej iteracji algorytmu, połowa ciał o większych masach jest stacjonarna (Wzór 2) i dobrana w pary z drugą połową ciał, które uderzają w nie z prędkością zależną od dystansu (Wzór 3) (Rys. 1).

$$v_i = 0, \quad i = 1, ...\frac{n}{2} \qquad \textit{(Wzór 2)}$$

$$v_i = x_i - x_{i\frac{n}{2}}, \quad i = \frac{n}{2} + 1, ...n \qquad \textit{(Wzór 3)}$$



Rysunek 1: Zderzanie par ciał

Po zderzeniu obliczane są nowe prędkości oraz położenia (Wzory 4-6).

$$v_i' = \frac{(m_{i+\frac{n}{2}} + \varepsilon m_{i+\frac{n}{2}})v_{i+\frac{n}{2}}}{m_i + m_{i+\frac{n}{2}}}, \quad x_i^{new} = x_i + rand \circ v_i', \quad i = 1, ..., \frac{n}{2} \qquad \textit{(Wzór 4)}$$

$$v_i' = \frac{(m_i - \varepsilon m_{i-\frac{n}{2}})v_i}{m_i + m_{i-\frac{n}{2}}}, \quad x_i^{new} = x_{i-\frac{n}{2}} + rand \circ v_i', \quad i = \frac{n}{2} + 1, ..., n \qquad \textit{(Wzór 5)}$$

$$\varepsilon = 1 - \frac{iter}{iter_{max}} \qquad \textit{(Wzór 6)}$$

Epsilon jest współczynnikiem odbicia z zakresu [0;1]. Uzasadnienie wzoru 6 znajduje się w oryginalnym artykule prezentującym algorytm.
Proces jest powtarzany do spełnienia warunku końcowego, zazwyczaj do pewnej ilości iteracji.
Położenia ciał są przestrzenią wyników, a szukanie nowych położeń i aktualizowanie mas oraz doboru stacjonarnych bloków pozwalają na znalezienie optymalnych rozwiązań w tej przestrzeni.

3. Kod źródłowy

Kod źródłowy znajduje się w repozytorium GitHub:
https://github.com/vis4rd/cim_project_2022

4. Realizacja projektu

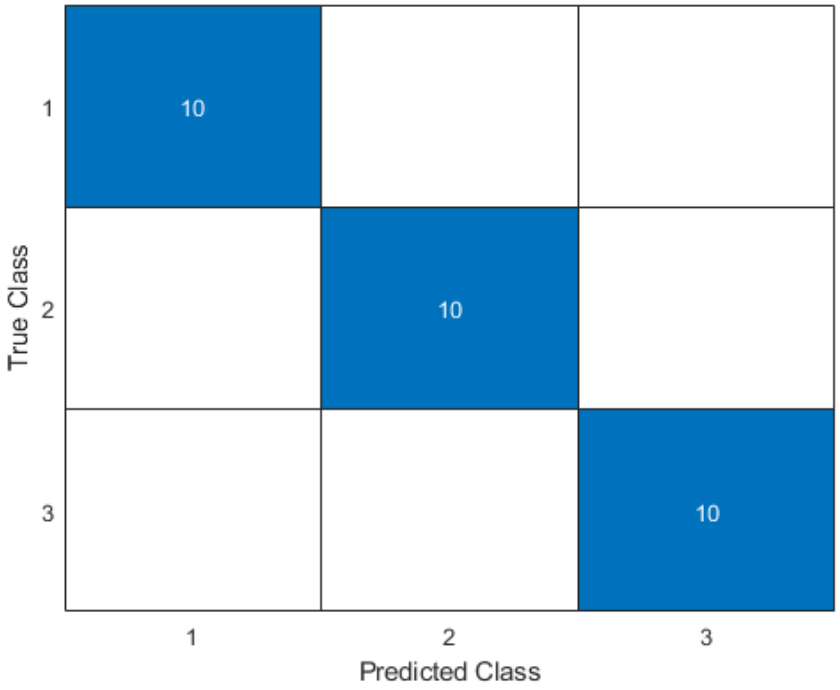
Projekt został zrealizowany w środowisku MATLAB.
Szkic programu:

1. Na początku wczytaliśmy dane z plików *iris.data*, *wine.data* oraz *seeds.data*.
2. Każdy zestaw danych dzielimy na dane uczące i testujące.
3. Do generowania FIS wybieramy opcję *'SubtractiveClustering'*.
4. Generujemy FIS za pomocą funkcji *genfis*.
5. Pobieramy parametry wygenerowanego FIS za pomocą funkcji *getTunableSettings*.
6. Ustalamy inne potrzebne wartości do optymalizacji parametrów, takie jak ilość parametrów oraz liczba iteracji algorytmu.
7. Optymalizujemy pierwotne parametry FIS za pomocą funkcji *Optimize*, która jest implementacją algorytmu CBO.
8. Ewaluujemy utworzony FIS na naszym zbiorze testowym.
9. Liczymy dokładność jako wartość średniokwadratową z różnicy pomiędzy wartościami uzyskanymi a oczekiwanymi.
10. Następnie dzielimy nasze dane uczące zgodnie z algorytmem *CV-5*.
11. Dla każdego folda tworzymy FIS, ewaluujemy go i liczymy naszą dokładność.
12. Końcową dokładność liczymy jako średnią arytmetyczną dokładności dla każdego folda.
13. Tworzymy macierz pomyłek.

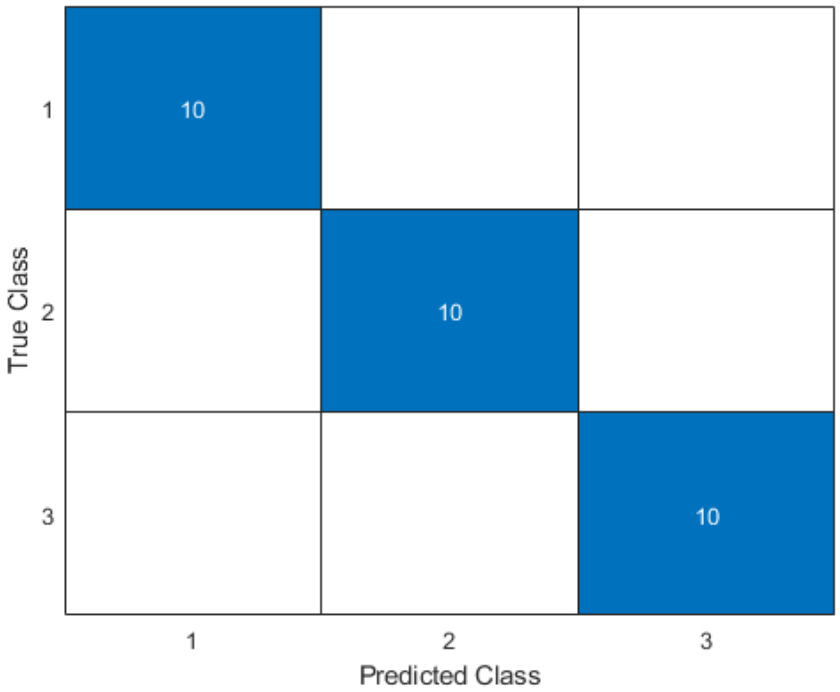
Dla zbioru WINE użyliśmy podczas generacji FIS użyliśmy wartości 1 dla opcji *ClusterInfluenceRange*. Umożliwiło nam to generację zbioru o mniejszej liczbie parametrów, w przeciwnym wypadku wykonywanie algorytmu trwało zbyt długo do terminowego zrealizowania projektu.

5. Wyniki

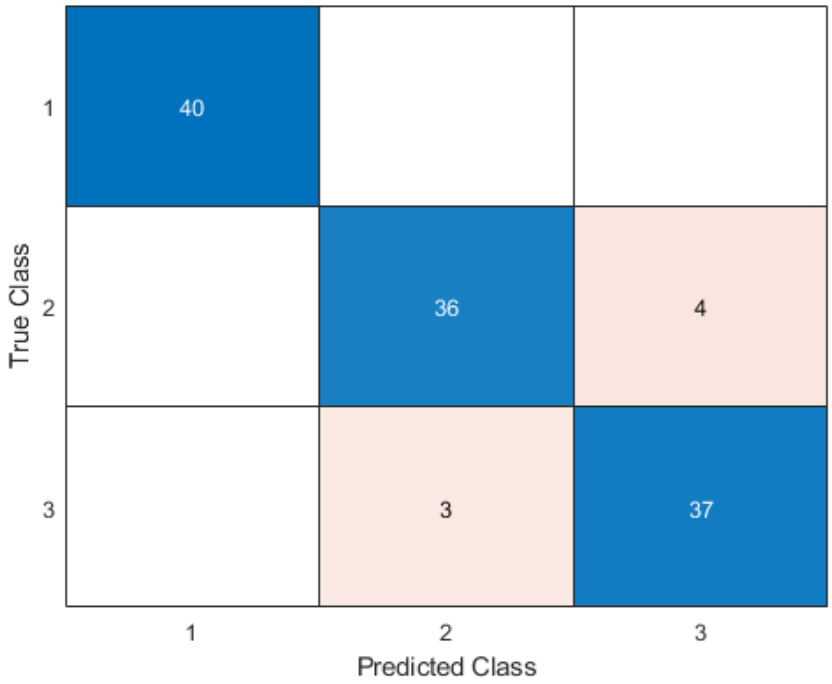
5.1. Zbiór IRIS



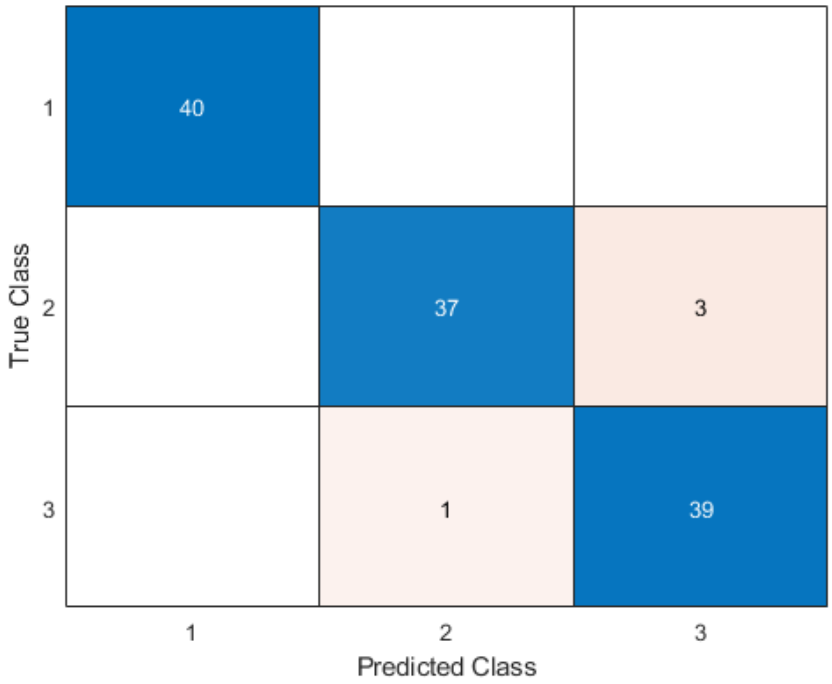
Rysunek 3: Macierz pomyłek IRIS przed zbioru testującego przed optymalizacją



Rysunek 4: Macierz pomyłek IRIS zbioru testującego po optymalizacji



Rysunek 5: Macierz pomyłek IRIS walidacji CV5 przed optymalizacją



Rysunek 6: Macierz pomyłek IRIS walidacji CV5 po optymalizacji

Czułość:

- zbiór testujący przed optymalizacją parametrów: 100%, 100%, 100%
- zbiór testujący po optymalizacji parametrów: 100%, 100%, 100%
- walidacja CV5 przed optymalizacją parametrów z użyciem CV5: 100%, 90%, 92.5%
- walidacja CV5 po optymalizacji parametrów z użyciem CV5: 100%, 92.5%, 97.5%

Swoistość:

- zbiór testujący przed optymalizacją parametrów: 100%, 100%, 100%
- zbiór testujący po optymalizacji parametrów: 100%, 100%, 100%
- walidacja CV5 przed optymalizacją parametrów z użyciem CV5: 100%, 96.25%, 95%
- walidacja CV5 po optymalizacji parametrów z użyciem CV5: 100%, 98.75%, 96.25%

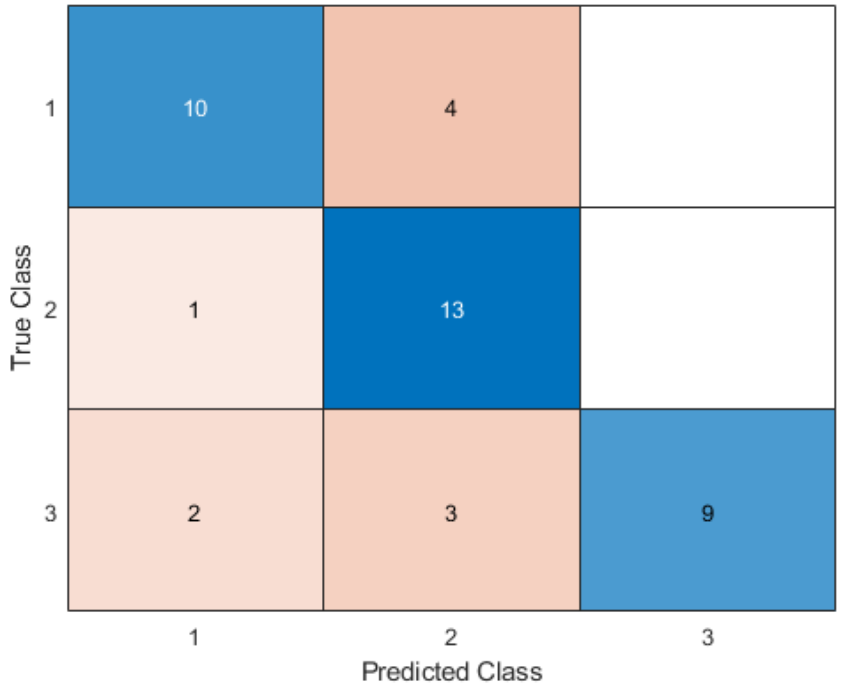
Precyzja:

- zbiór testujący przed optymalizacją parametrów: 100%, 100%, 100%
- zbiór testujący po optymalizacji parametrów: 100%, 100%, 100%
- walidacja CV5 przed optymalizacją parametrów z użyciem CV5: 100%, 92.3077%, 90.2439%
- walidacja CV5 po optymalizacji parametrów z użyciem CV5: 100%, 97.3684%, 92.8571%

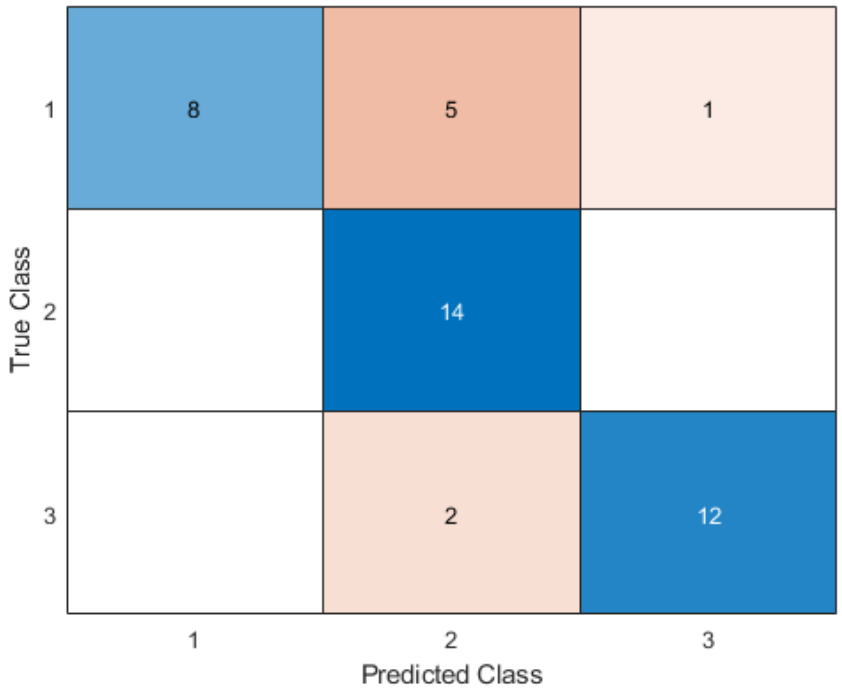
Dokładność:

- zbiór testujący przed optymalizacją parametrów: 98.6527%
- zbiór testujący po optymalizacji parametrów: 99.0757%
- walidacja CV5 przed optymalizacją parametrów z użyciem CV5: 94.1667%
- walidacja CV5 po optymalizacji parametrów z użyciem CV5: 96.6667%

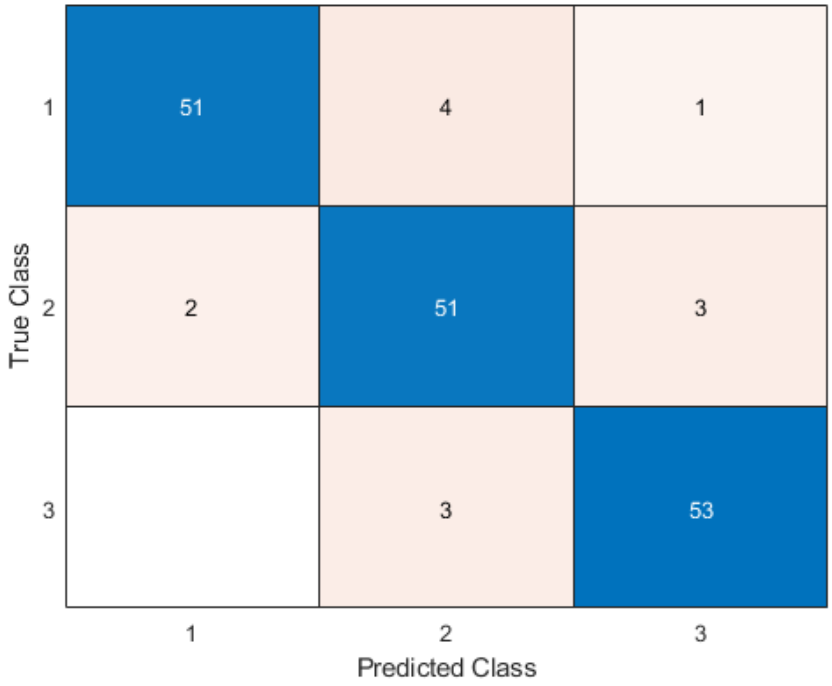
5.2. Zbiór SEEDS



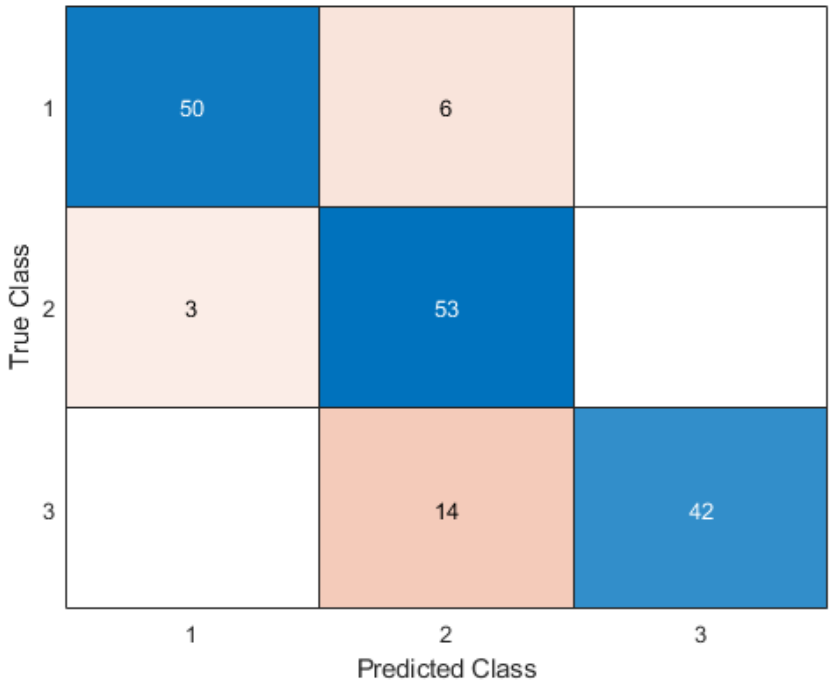
Rysunek 7: Macierz pomyłek SEEDS przed zbioru testującego przed optymalizacją



Rysunek 8: Macierz pomyłek SEEDS zbioru testującego po optymalizacji



Rysunek 9: Macierz pomyłek SEEDS walidacji CV5 przed optymalizacją



Rysunek 10: Macierz pomyłek SEEDS walidacji CV5 po optymalizacji

Czułość:

- zbiór testujący przed optymalizacją parametrów: 71.4286%, 92.8571%, 64.2857%
- zbiór testujący po optymalizacji parametrów: 57.1429%, 100%, 85.71429%
- walidacja CV5 przed optymalizacją parametrów z użyciem CV5: 91.0714%, 91.0714%, 94.6429%
- walidacja CV5 po optymalizacji parametrów z użyciem CV5: 89.2857%, 94.6429%, 75%

Swoistość:

- zbiór testujący przed optymalizacją parametrów: 89.2857%, 75%, 100%
- zbiór testujący po optymalizacji parametrów: 100%, 75%, 96.4286%
- walidacja CV5 przed optymalizacją parametrów z użyciem CV5: 98.2143%, 93.75%, 96.4286%
- walidacja CV5 po optymalizacji parametrów z użyciem CV5: 97.3214%, 82.1429%, 100%

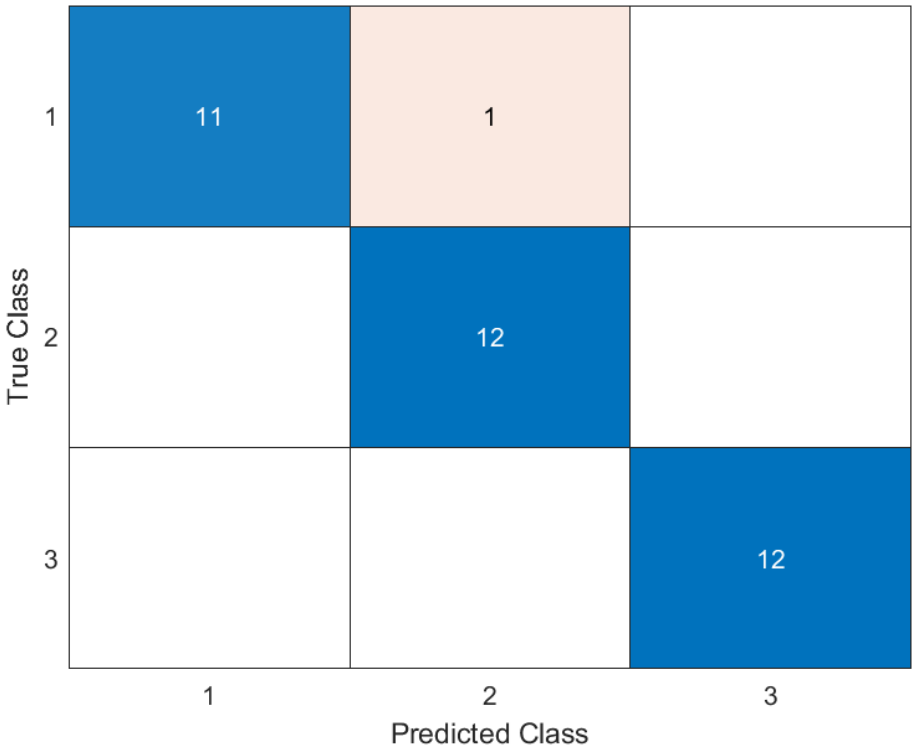
Precyzja:

- zbiór testujący przed optymalizacją parametrów: 76.9231%, 65%, 100%
- zbiór testujący po optymalizacji parametrów: 100%, 66.6667%, 92.3077%
- walidacja CV5 przed optymalizacją parametrów z użyciem CV5: 96.2264%, 87.931%, 92.9825%
- walidacja CV5 po optymalizacji parametrów z użyciem CV5: 94.3396%, 72.6027%, 100%

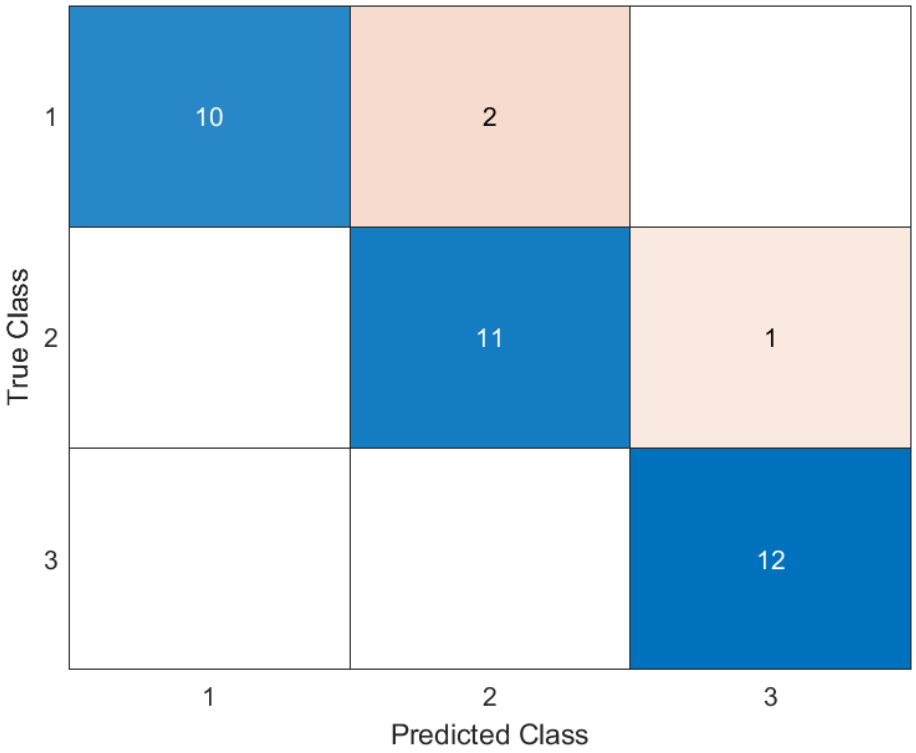
Dokładność:

- zbiór testujący przed optymalizacją parametrów: 59.0432%
- zbiór testujący po optymalizacji parametrów: 74.5454%
- walidacja CV5 przed optymalizacją parametrów z użyciem CV5: 90.4565%
- walidacja CV5 po optymalizacji parametrów z użyciem CV5: 86.2002%

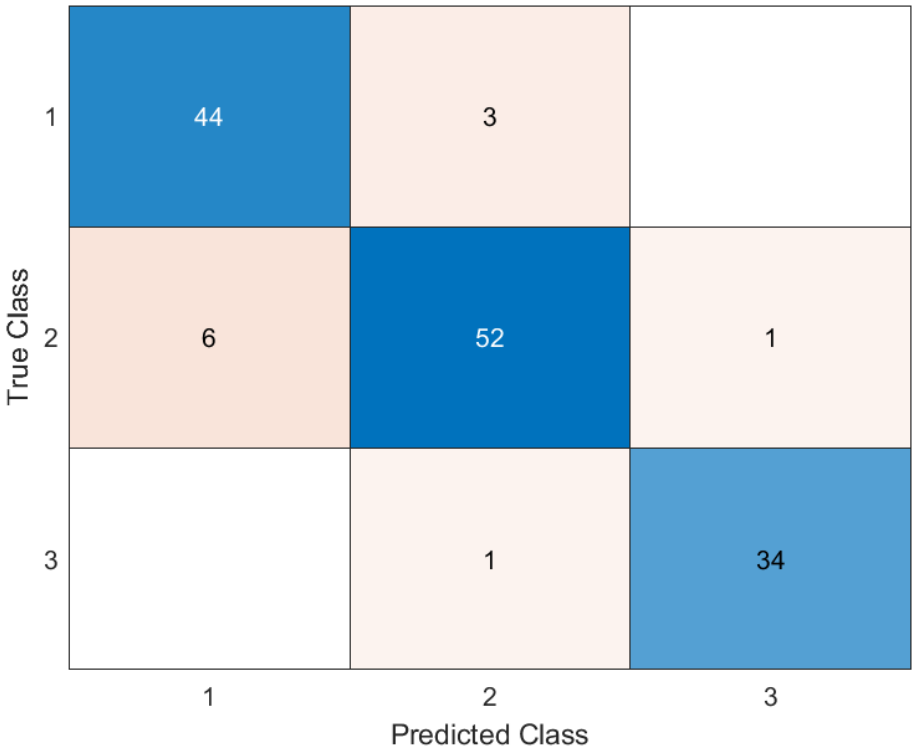
5.3. Zbiór WINE



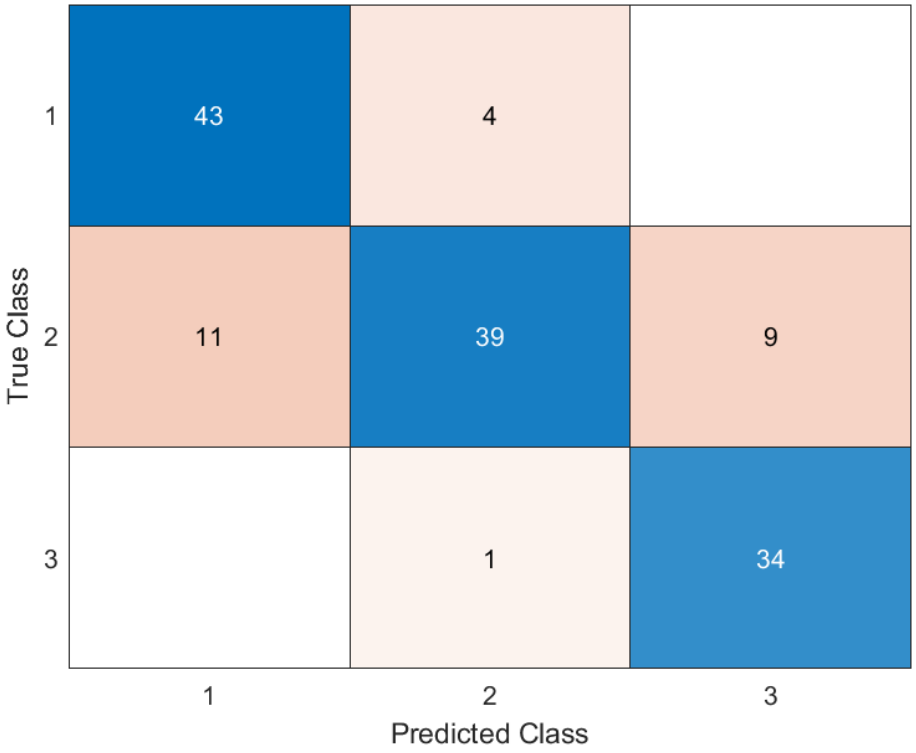
Rysunek 11: Macierz pomyłek WINE przed zbioru testującego przed optymalizacją



Rysunek 12: Macierz pomyłek WINE zbioru testującego po optymalizacji



Rysunek 13: Macierz pomyłek WINE walidacji CV5 przed optymalizacją



Rysunek 14: Macierz pomyłek WINE walidacji CV5 po optymalizacji

Czułość:

- zbiór testujący przed optymalizacją parametrów: 91.6667%, 100%, 100%
- zbiór testujący po optymalizacji parametrów: 83.3333%, 91.6667%, 100%
- walidacja CV5 przed optymalizacją parametrów z użyciem CV5: 93.617%, 88.1356%, 97.1429%
- walidacja CV5 po optymalizacji parametrów z użyciem CV5: 91.4894%, 66.1017%, 97.1429%

Swoistość:

- zbiór testujący przed optymalizacją parametrów: 100%, 95.8333%, 100%
- zbiór testujący po optymalizacji parametrów: 100%, 91.6667%, 95.8333%
- walidacja CV5 przed optymalizacją parametrów z użyciem CV5: 93.617%, 95.122%, 99.0566%
- walidacja CV5 po optymalizacji parametrów z użyciem CV5: 88.2979%, 93.9024%, 91.5094%

Precyzja:

- zbiór testujący przed optymalizacją parametrów: 100%, 92.3077%, 100%
- zbiór testujący po optymalizacji parametrów: 100%, 84.6154%, 92.3077%
- walidacja CV5 przed optymalizacją parametrów z użyciem CV5: 88%, 92.8571%, 97.1429%
- walidacja CV5 po optymalizacji parametrów z użyciem CV5: 79.6296%, 88.6364%, 79.0698%

Dokładność:

- zbiór testujący przed optymalizacją parametrów: 93.5688%
- zbiór testujący po optymalizacji parametrów: 91.1024%
- walidacja CV5 przed optymalizacją parametrów z użyciem CV5: 92.2131%
- walidacja CV5 po optymalizacji parametrów z użyciem CV5: 82.2323%

6. Wnioski

Działanie algorytmu optymalizacji CBO najlepiej widać na zbiorze SEEDS. Po nauczeniu układu Fuzzy Logic, klasyfikacja danych testujących wskazuje gorsze wyniki dla tego zbioru (dokładność 59%). Po zastosowaniu optymalizacji, dokładność poprawia się o około 15 punktów procentowych.

Dokładności dla zbiorów IRIS oraz WINE w układach generowanych ogólną zasadą są stosunkowo wysokie. Z tego powodu może się zdarzyć, że po optymalizacji klasyfikacja zbioru testowego wcale nie będzie lepsza. Przyczyną może być overfitting klasyfikatora dla zbioru uczącego.

Jak można zauważyć na rysunkach 4. oraz 5., skuteczność klasyfikacji jest na tyle wysoka, że zbiór testujący poprawnie wskazuje wynik za każdym razem. Wynika to z

faktu, że zbiór testujący jest niewielki, a dane wejściowe są do siebie podobne na tyle, że przy zaokrąglaniu wyników do przedstawienia na macierzy pomyłek nie pojawia się ani jeden definitywny błąd klasyfikacji. Jednak do obliczenia dokładności zostały użyte różnice nie zaokrąglonych danych.

7. Bibliografia

Kaveh, A., Mahdavi, V. (2014). Colliding bodies optimization: A novel meta-heuristic method. *Computers & Structures*, *139*, 18-27. doi: 10.1016/j.compstruc.2014.04.005

Kaveh, A., Mahdavi, V. (2015). *Colliding Bodies Optimization: Extensions and Applications*. Cham: Springer. DOI: 10.1007/978-3-319-19659-6

http://scikit-learn.org/stable/modules/cross_validation.html

Zestawy danych:

- http://archive.ics.uci.edu/ml/datasets/Iris
- http://archive.ics.uci.edu/ml/datasets/seeds
- http://archive.ics.uci.edu/ml/datasets/Wine