

Zadanie nr 3 - Detekcja anomalii w zbiorze danych

Analiza danych złożonych z detekcją wyjątków

Karol Kazusek - 254189, Sebastian Zych - 254264

12.11.2024

1 Cel zadania

Zadanie polega na detekcji anomalii w wybranym strumieniu danych. Do analiz należało wybrać tematykę medycyny lub rozpoznawania faz ruchu aktywności człowieka (systemy HAR)

2 Opis zaproponowanych algorytmów detektorów anomalii

2.1 Metody detekcji

2.1.1 DBSCAN

Metoda wykorzystująca **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)** to algorytm klastrowania, który jest szczególnie przydatny do wykrywania anomalii w zbiorach danych. Jego główną zaletą jest to, że może skutecznie identyfikować gęsto zamieszkane obszary (klastry) w danych oraz odróżniać je od obszarów rzadkich, które uznaje się za "szum" lub anomalie. Dzięki analizie gęstości, DBSCAN może odróżnić obszary z dużą liczbą punktów (klastry) od obszarów z niską gęstością, które wskazują na anomalie. DBSCAN wykorzystuje dwa kluczowe parametry:

- **Epsilon ϵ** : maksymalna odległość, w której punkty są uznawane za należące do tego samego klastra.
- **min sample**: minimalna liczba punktów, które muszą znajdować się w otoczeniu o promieniu ϵ , aby obszar mógł zostać uznany za klaster.

2.1.2 LOF

LOF (Local Outlier Factor) to metoda wykrywania anomalii, która opiera się na analizie **lokalnej gęstości** punktów w zbiorze danych. W przeciwieństwie do metod globalnych, LOF ocenia odstępstwa każdego punktu względem lokalnego otoczenia. To sprawia, że LOF jest szczególnie przydatny, gdy rozkład danych jest zróżnicowany i zawiera regiony o różnej gęstości.

2.1.3 COF

COF (Connectivity-Based Outlier Factor) to metoda detekcji anomalii, która, podobnie jak LOF, analizuje lokalne właściwości danych. COF jest szczególnie użyteczny, gdy dane są niejednorodne i mają różne gęstości, ale kluczową różnicą jest sposób, w jaki ocenia lokalną "odmienność" każdego punktu. COF bazuje na analizie "połączeń" pomiędzy punktami – czyli na analizie drogi, jaką trzeba przebyć, aby dotrzeć do sąsiadów danego punktu.

2.1.4 LUNAR

LUNAR (Local Unsupervised Noise and Anomaly detection with Randomized score) to nowoczesna metoda wykrywania anomalii, która bazuje na analizie lokalnej struktury danych. Algorytm jest stosowany w detekcji wartości odstających w złożonych i wielowymiarowych zbiorach danych, przy czym korzysta z podejścia losowego do oceny anomalii. Algorytm agreguje wyniki anomalii uzyskane z różnych losowych próbek, tworząc końcowy wynik dla każdego punktu. Na podstawie wartości końcowych LUNAR klasyfikuje punkty o wysokich ocenach jako anomalie.

3 Charakterystyka wybranych do danych

Anomaly Detection Falling People. W tym przypadku wybraliśmy dane HAR, dane zostawały uzyskiwane z 4 czujników założonych na:

1. klatkce piersiowej.
2. Pasku.
3. Oby kostkach.

Badania te miały umożliwić wykrywanie upadków osób starszych. Mierzyły one odległości w trzech wymiarach. W celu łatwej reprezentacji stworzyłem z odległości X i Y jedną miarę liczącą odległość od punktu zerowego, z myślą że ludzie upadają w danym kierunku.

Liczba próbek: 5804

Liczba kategorii: 2

Częstotliwość próbkowania: 10 Hz

4 Eksperymenty i wyniki

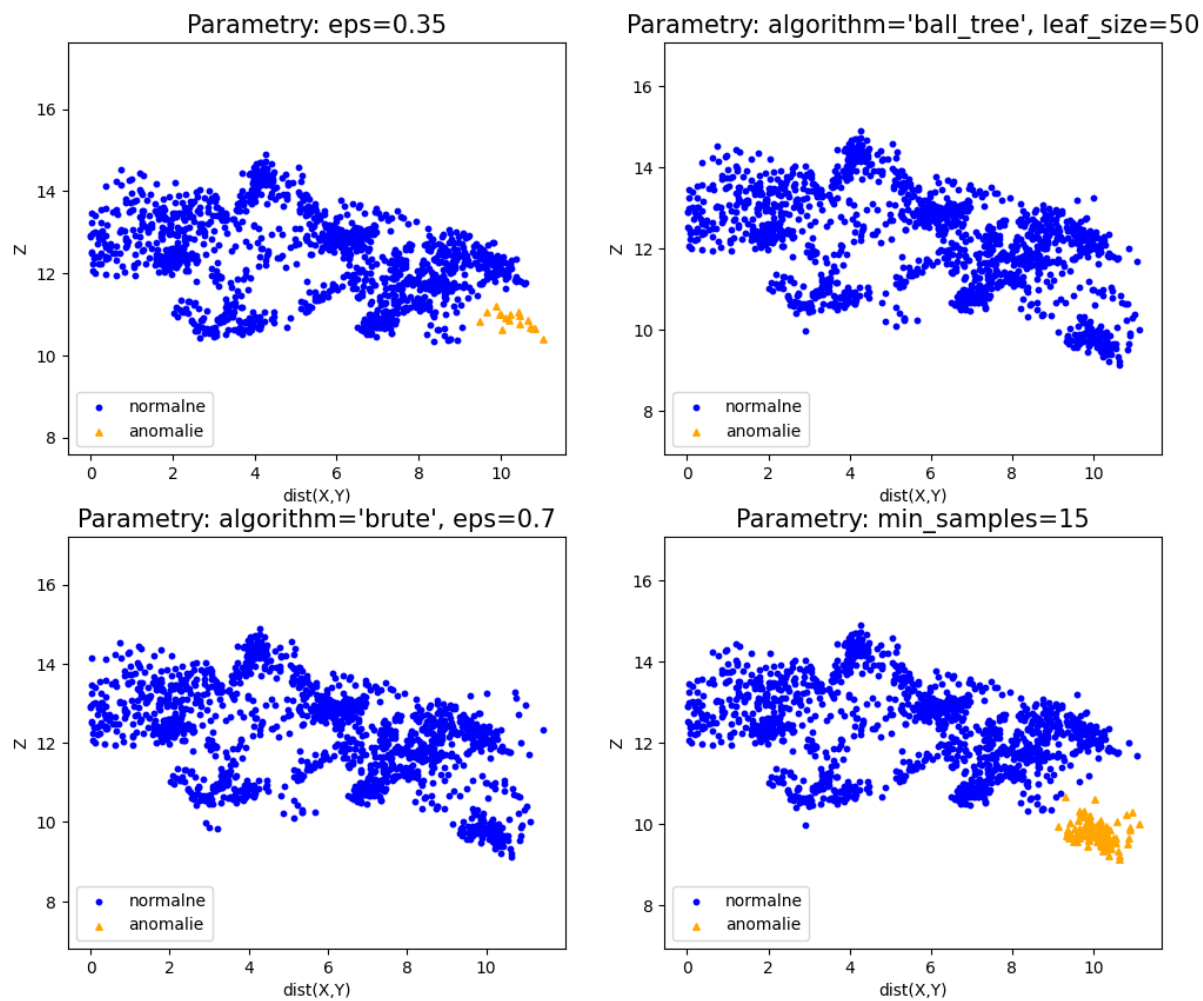
4.1 Eksperyment nr 1

4.1.1 Założenia

Wykonanie detekcji anomalii dla wybranych zbiorów danych obejmujące dane z różnych położonych sensorów.

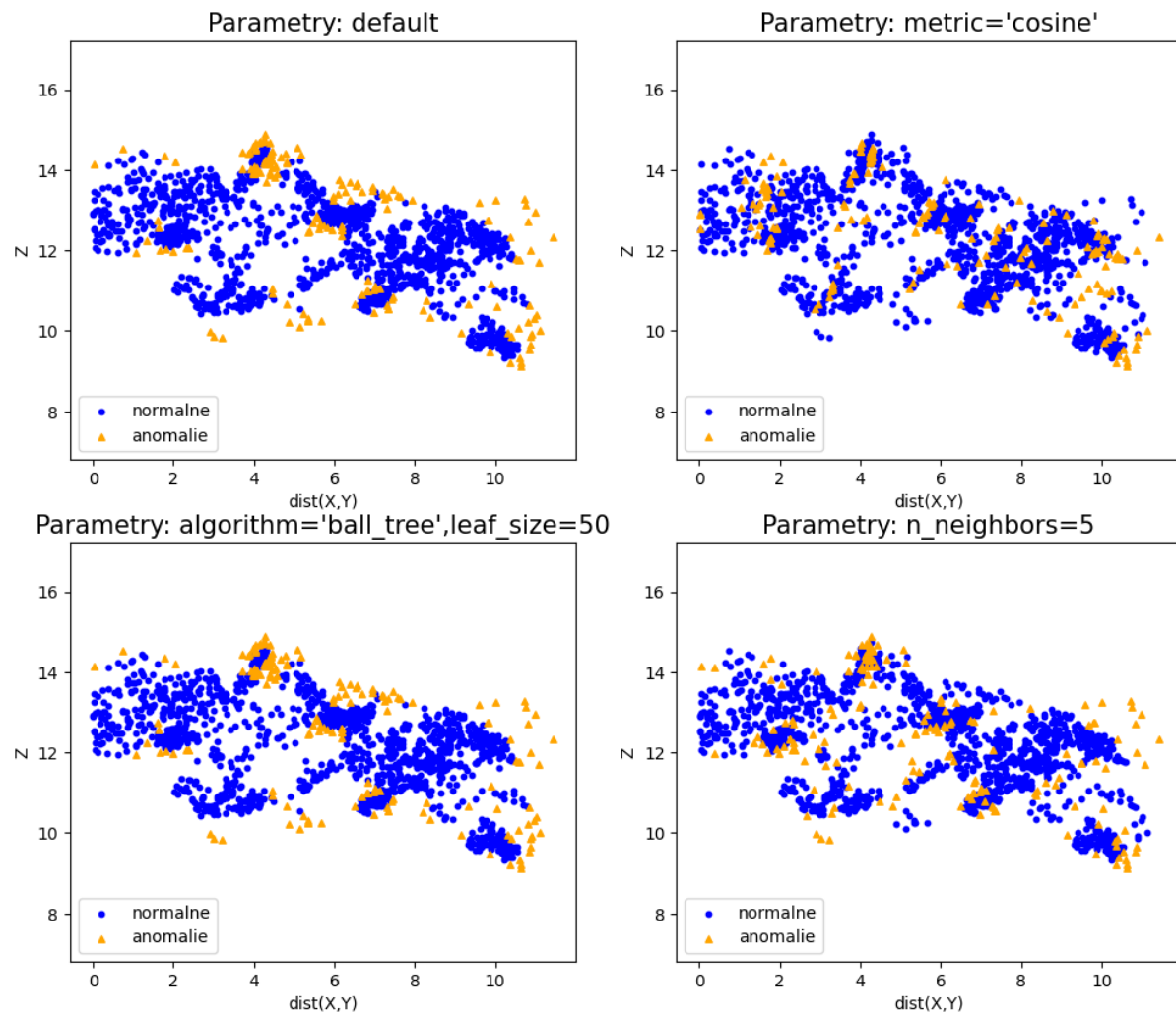
4.1.2 Wyniki dla sensora-1:

Detektor anomalii DBSCAN dla sensora 010-000-024-033



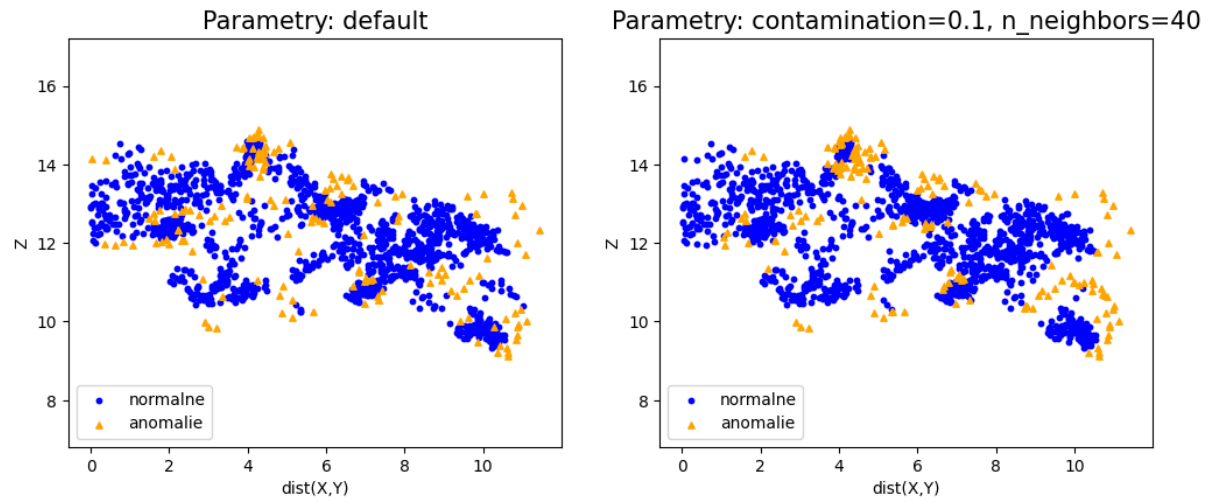
Rysunek 1: Wyniki dla DBSCAN sensor 1

Detektor anomalii LOF dla sensora 010-000-024-033



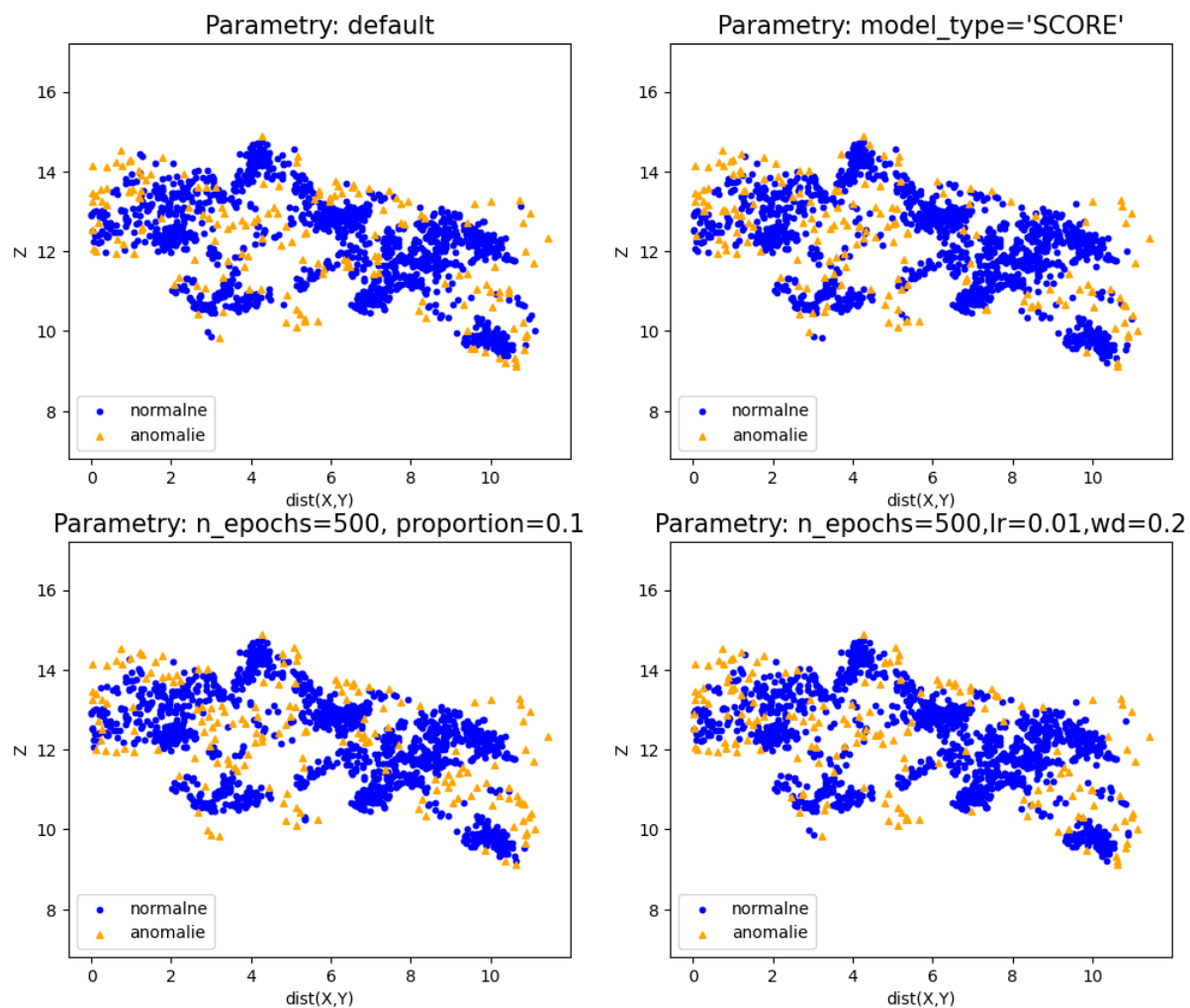
Rysunek 2: Wyniki dla LOF sensor 1

Detektor anomalii COF dla sensora 010-000-024-033



Rysunek 3: Wyniki dla COF sensor 1

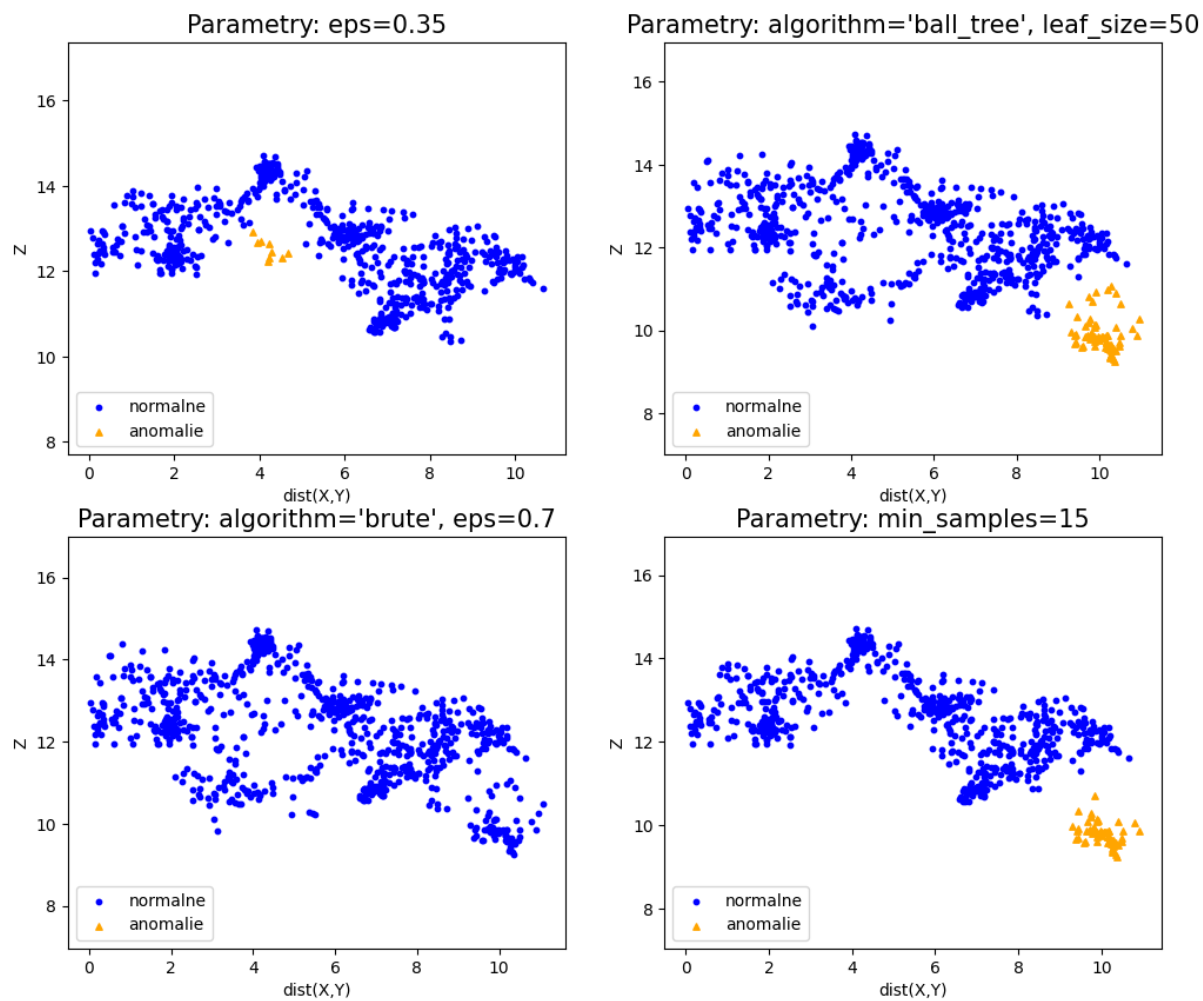
Detektor anomalii LUNAR dla sensora 010-000-024-033



Rysunek 4: Wyniki dla LUNAR sensor 1

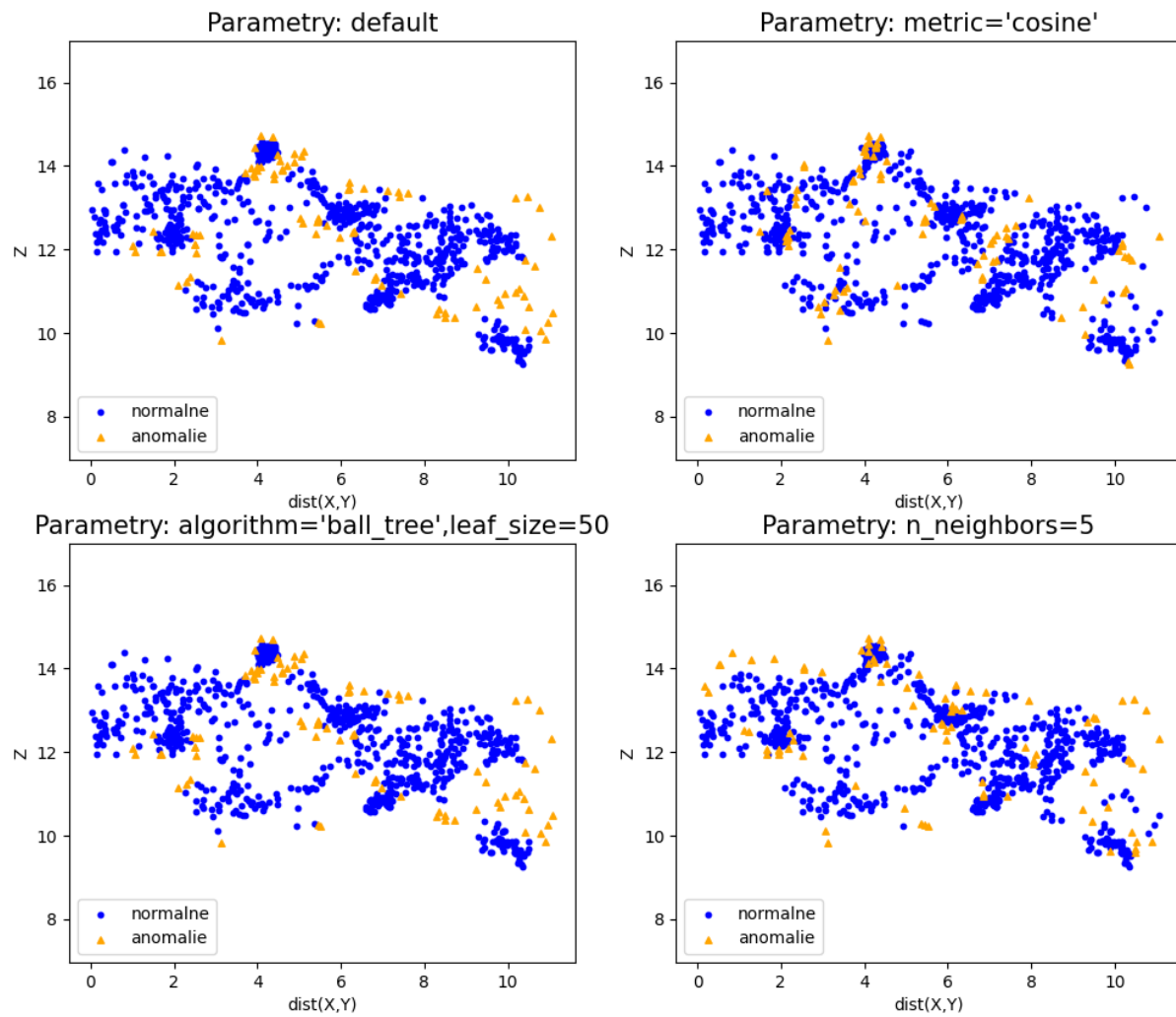
4.1.3 Wyniki dla sensora-2:

Detektor anomalii DBSCAN dla sensora 010-000-030-096



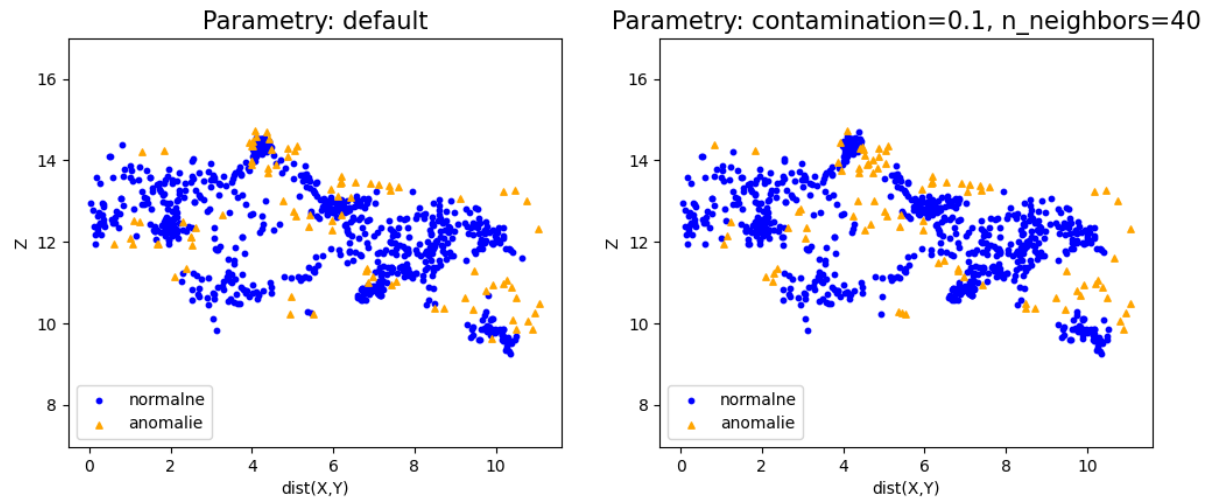
Rysunek 5: Wyniki dla DBSCAN sensor 2

Detektor anomalii LOF dla sensora 010-000-030-096



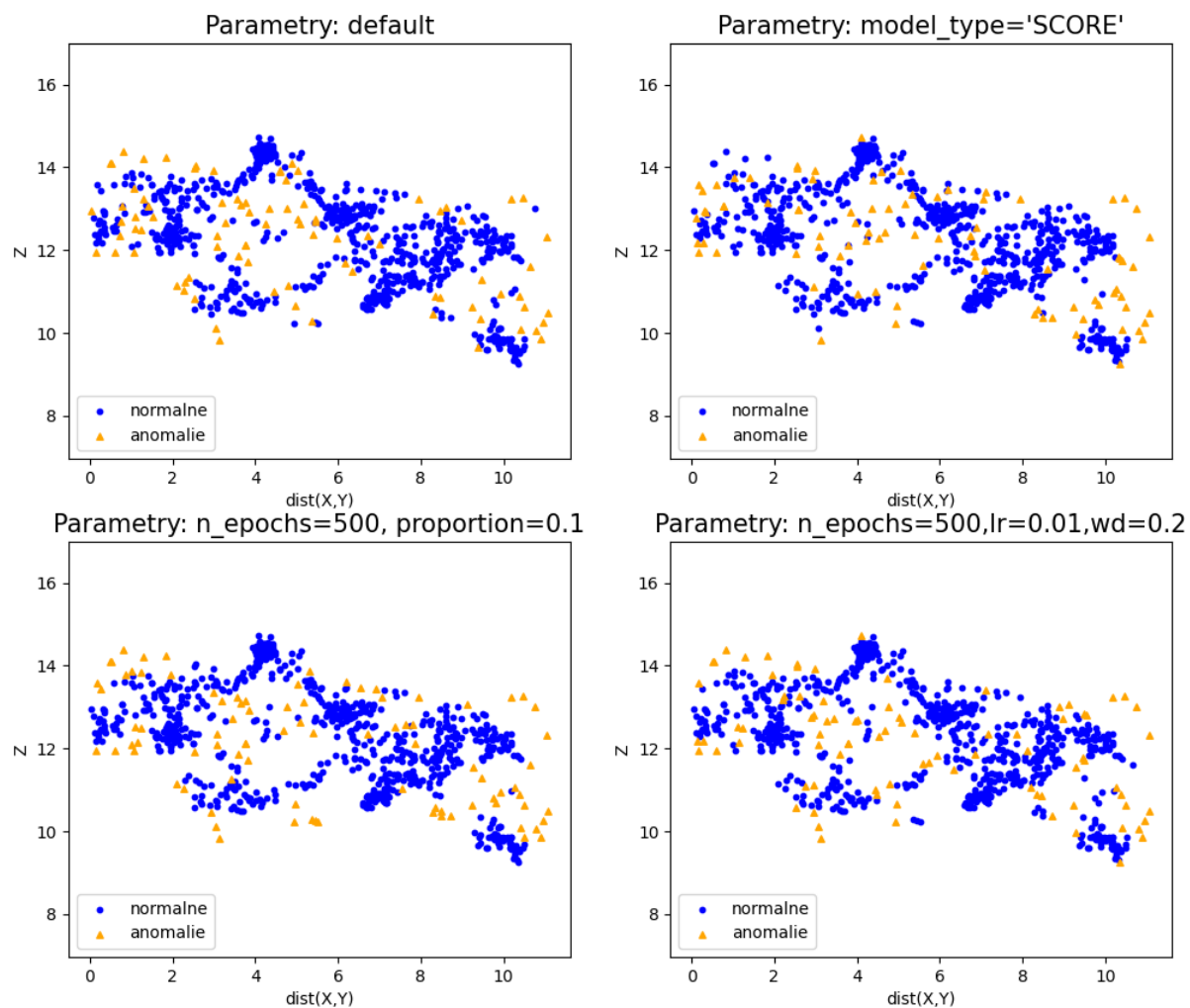
Rysunek 6: Wyniki dla LOF sensor 2

Detektor anomalii COF dla sensora 010-000-030-096



Rysunek 7: Wyniki dla COF sensor 2

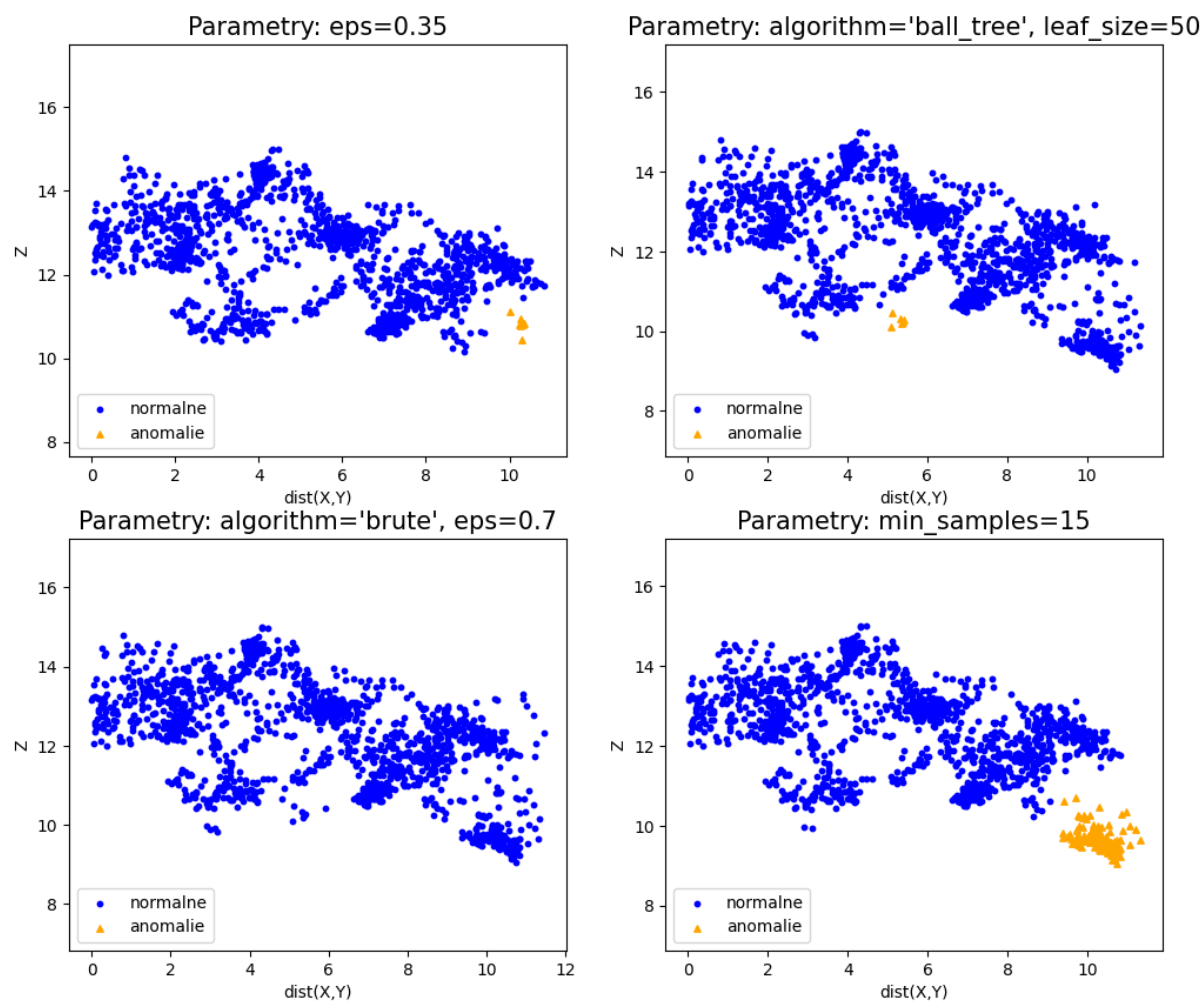
Detektor anomalii LUNAR dla sensora 010-000-030-096



Rysunek 8: Wyniki dla LUNAR sensor 2

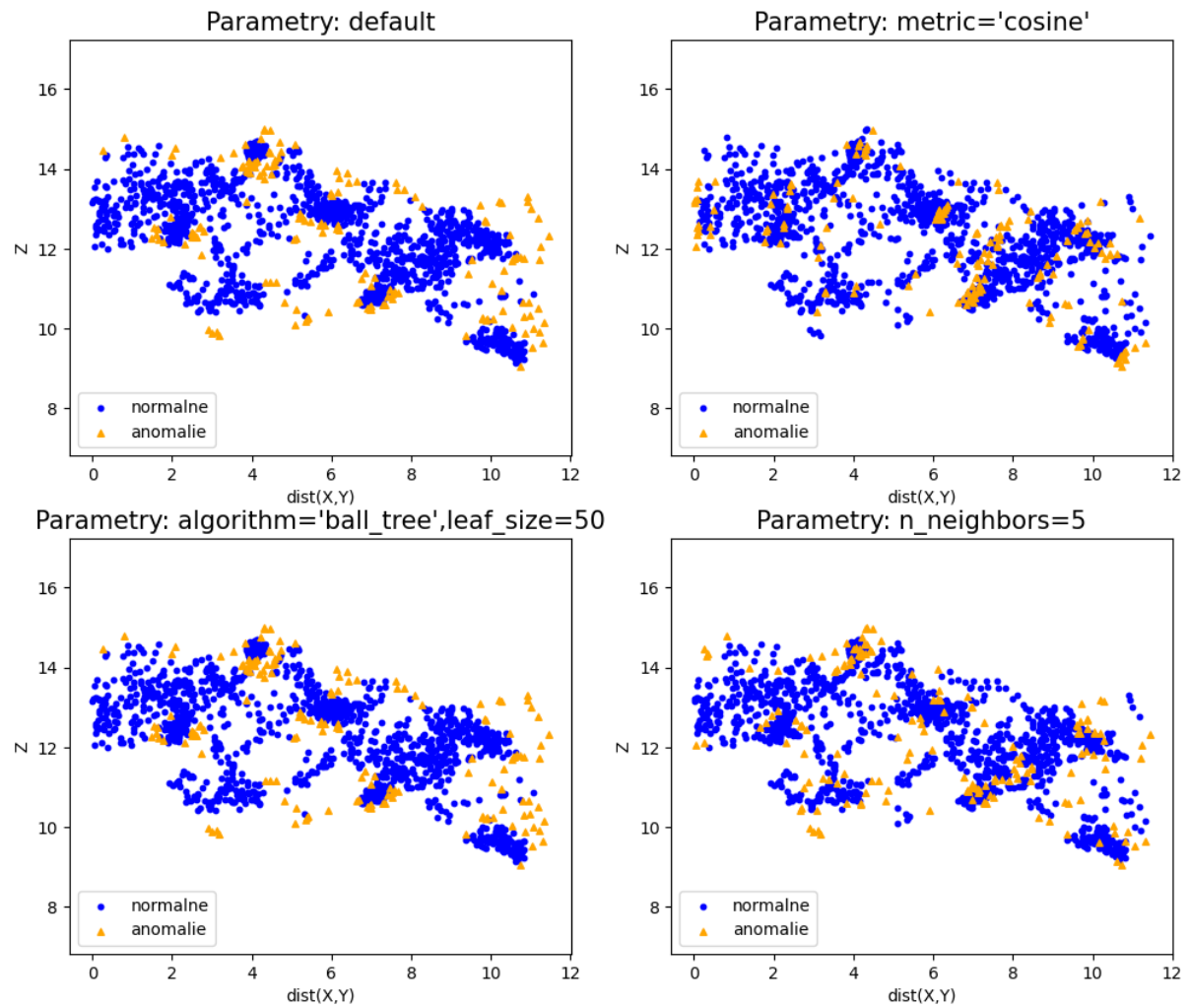
4.1.4 Wyniki dla sensora-3:

Detektor anomalii DBSCAN dla sensora 020-000-032-221



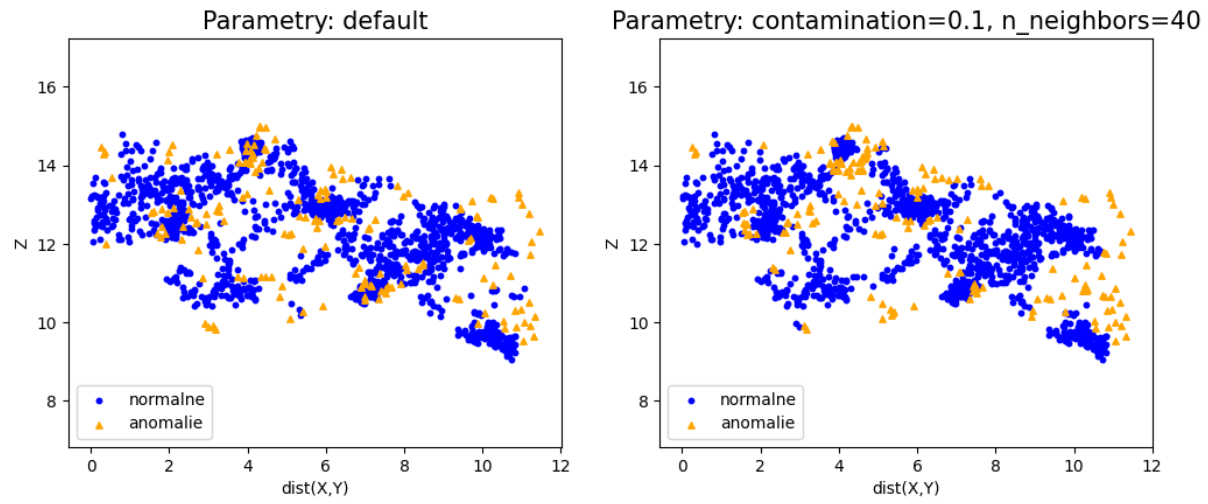
Rysunek 9: Wyniki dla DBSCAN sensor 3

Detektor anomalii LOF dla sensora 020-000-032-221



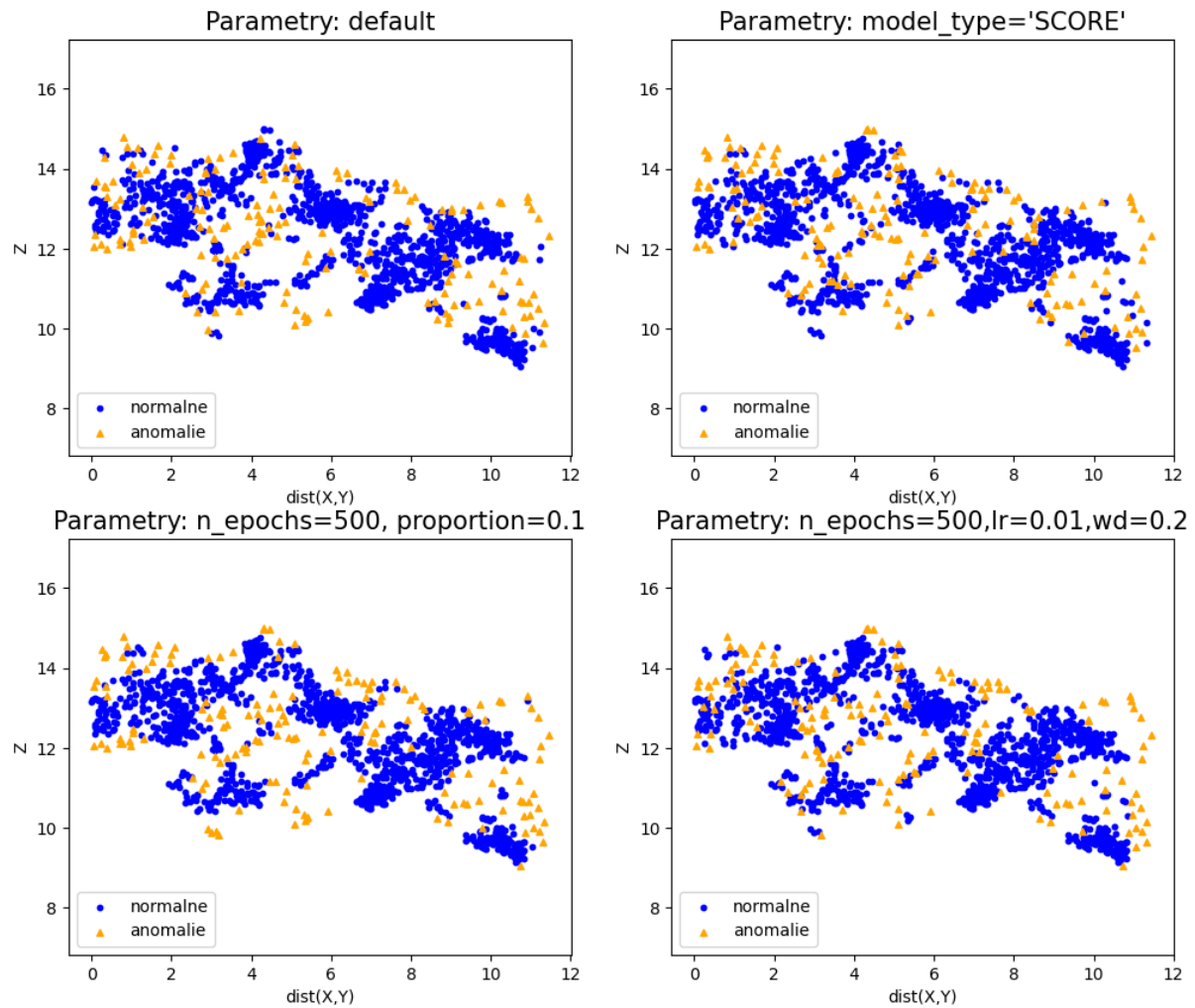
Rysunek 10: Wyniki dla LOF sensor 3

Detektor anomalii COF dla sensora 020-000-032-221



Rysunek 11: Wyniki dla COF sensor 3

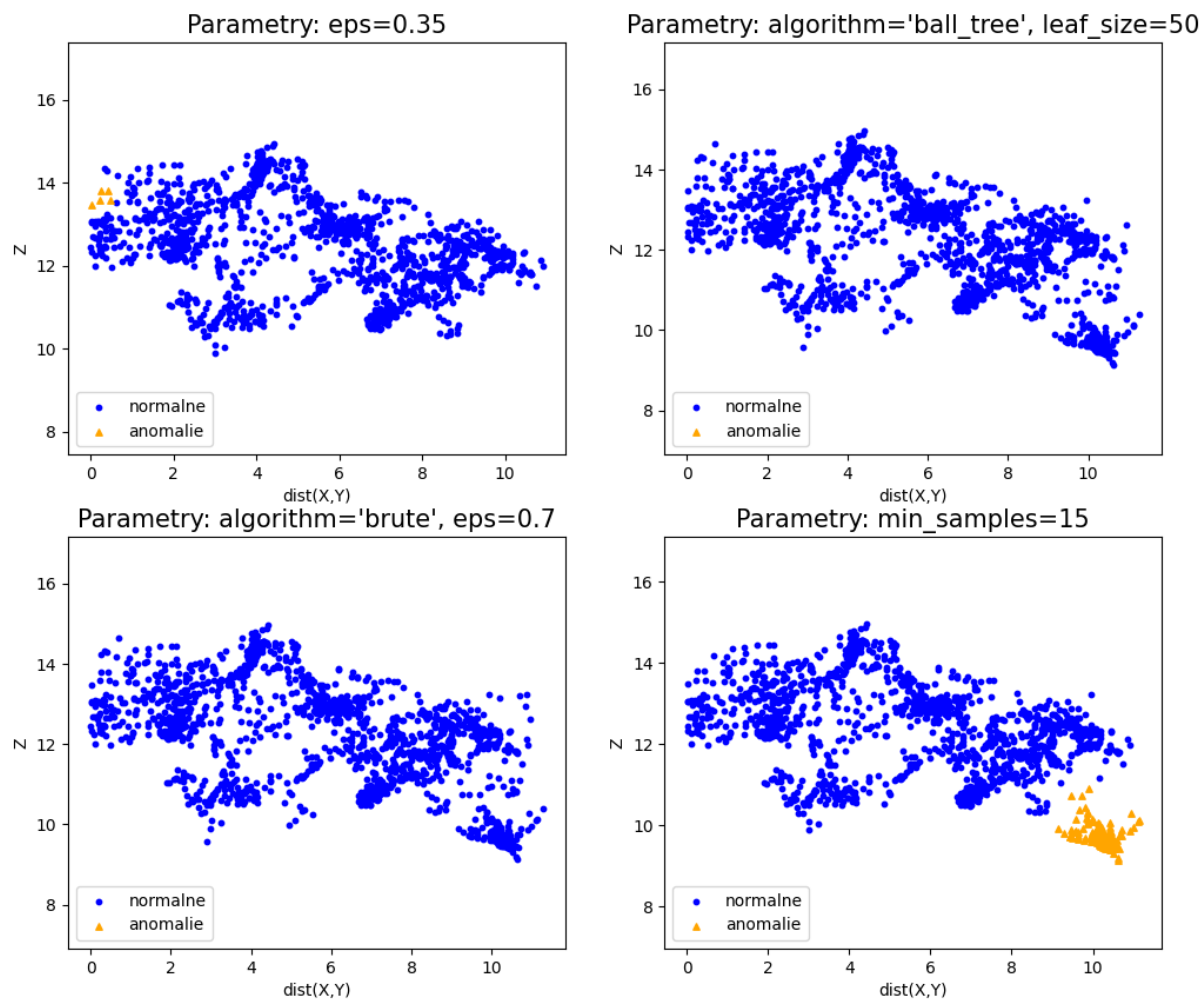
Detektor anomalii LUNAR dla sensora 020-000-032-221



Rysunek 12: Wyniki dla LUNAR sensor 3

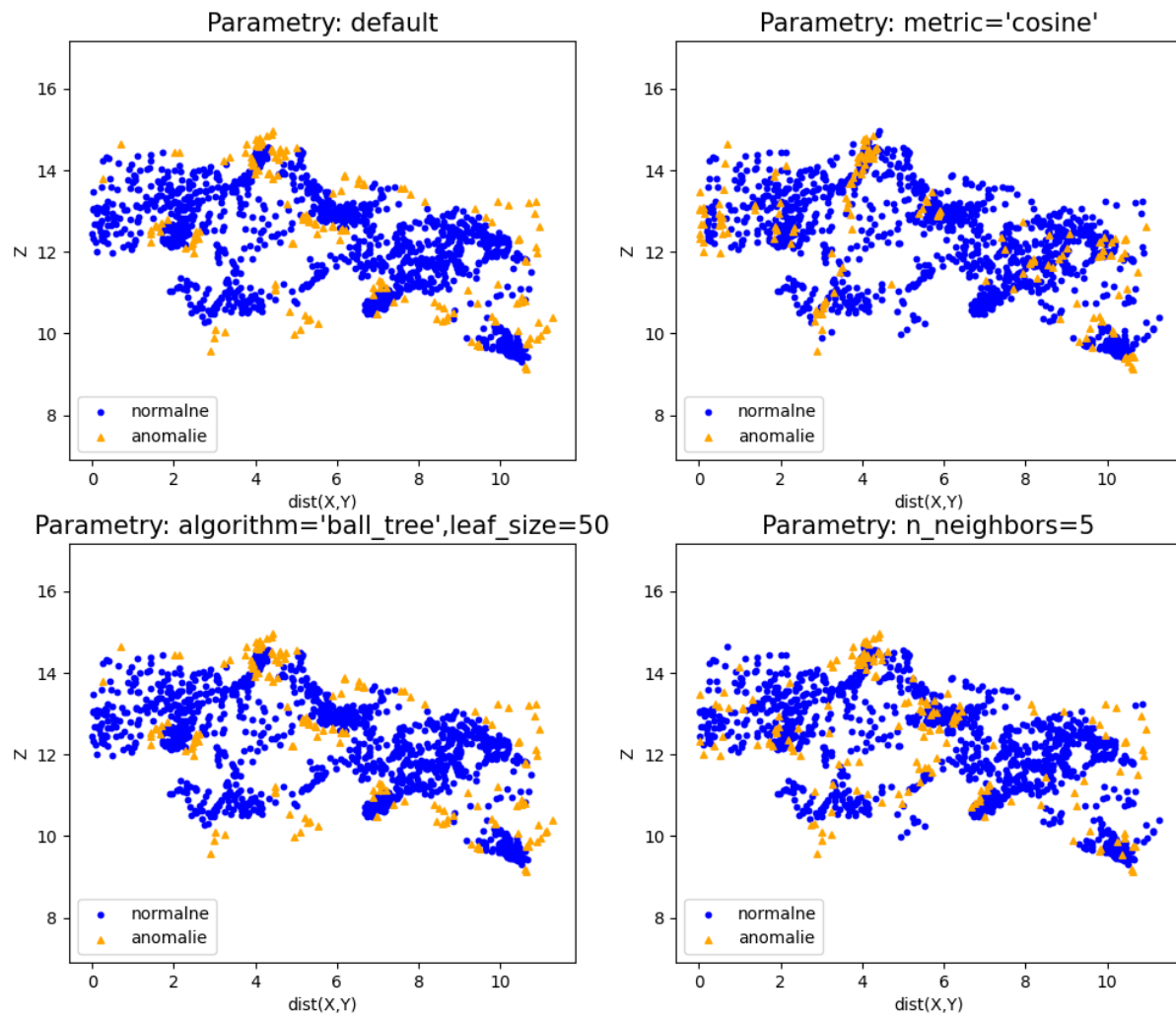
4.1.5 Wyniki dla sensora-4:

Detektor anomalii DBSCAN dla sensora 020-000-033-111



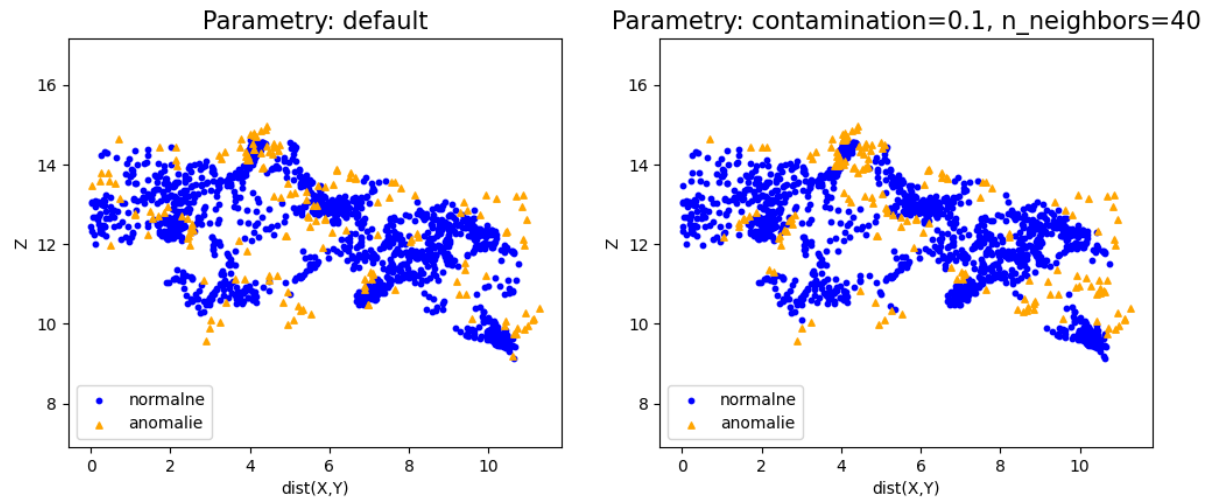
Rysunek 13: Wyniki dla DBSCAN sensor 4

Detektor anomalii LOF dla sensora 020-000-033-111



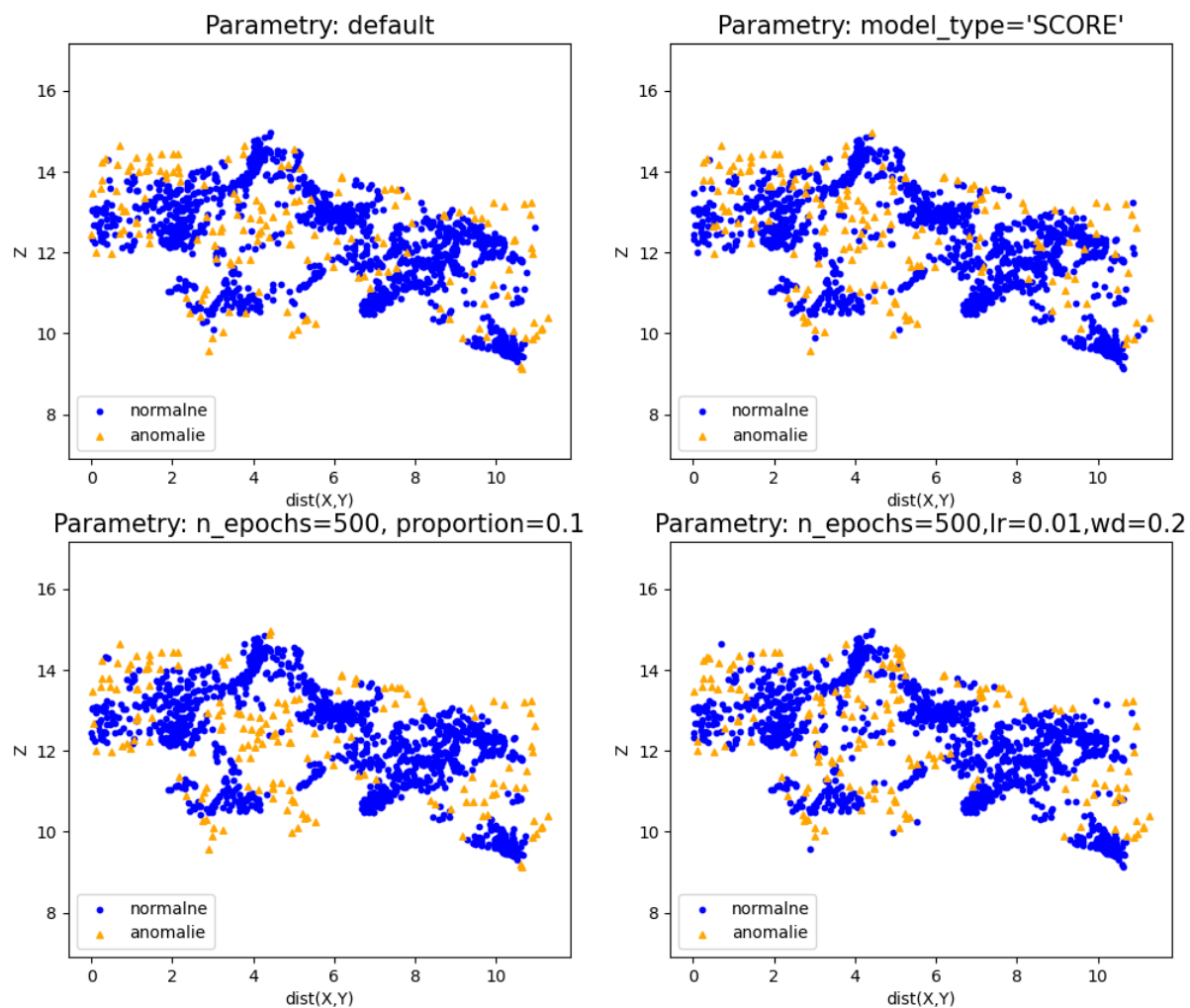
Rysunek 14: Wyniki dla LOF sensor 4

Detektor anomalii COF dla sensora 020-000-033-111



Rysunek 15: Wyniki dla COF sensor 4

Detektor anomalii LUNAR dla sensora 020-000-033-111



Rysunek 16: Wyniki dla LUNAR sensor 4

5 Wnioski

Wnioski z przeprowadzonych eksperymentów dowodzą, że

- DBSCAN jako jedyny detektor z poprawnymi parametrami wykrył anomalie jako oddzielną klasę upadków. Było to możliwe dzięki sposobie tworzenia klastrów. Wykorzystując globalną gęstość, zamiast lokanych jak w pozostałych metodach.
- Pozostałe detektory wykrywały lokalne anomalie. Należy zauważyć że metoda lunar często pomija niektóre anomalie, zostawiając 1,2 elementy jako zbiór normalnych pomiarów gdzie LOF oraz COF przypisują tym rekordą etykiety anomalii.