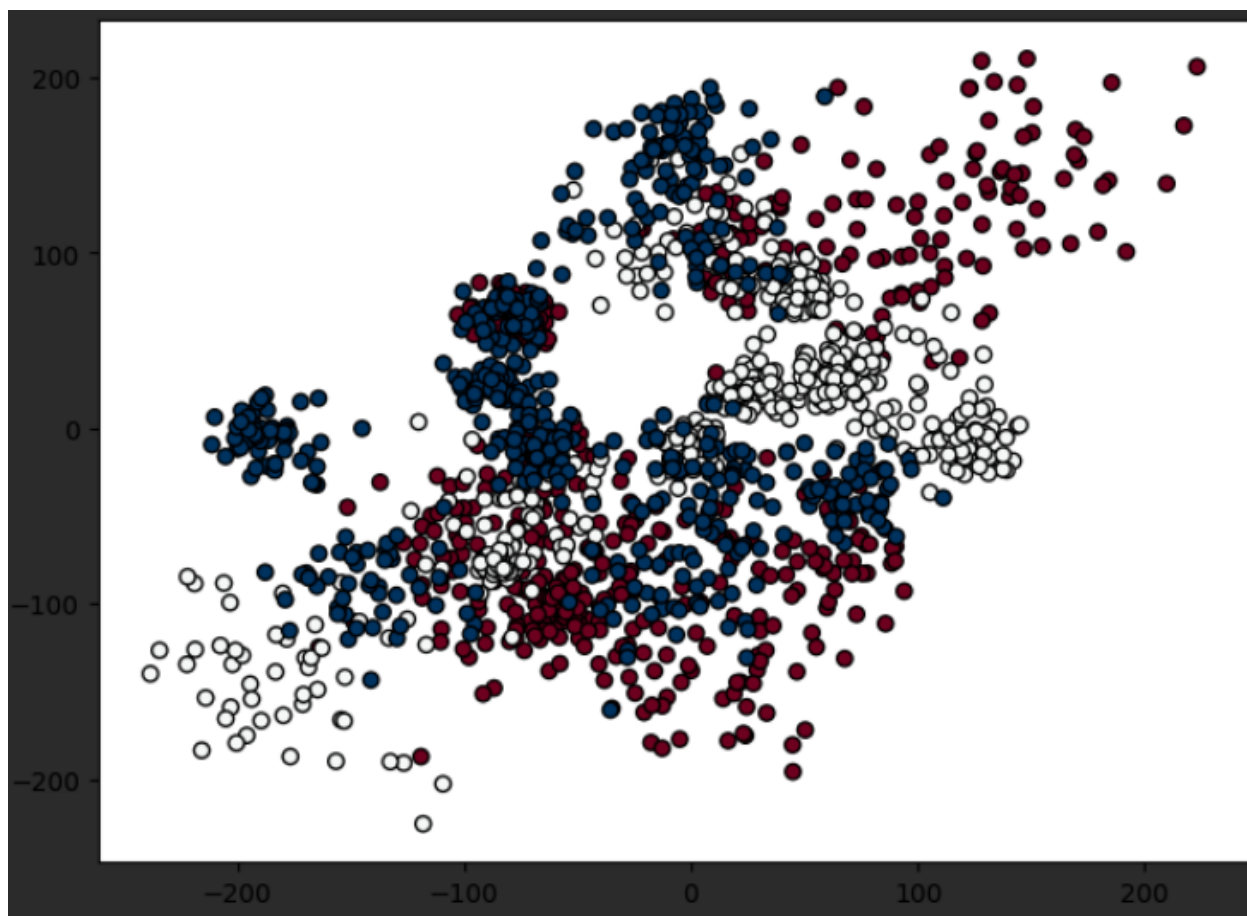


Podstawy Uczenia Maszynowego - Eksperymenty z k-NN

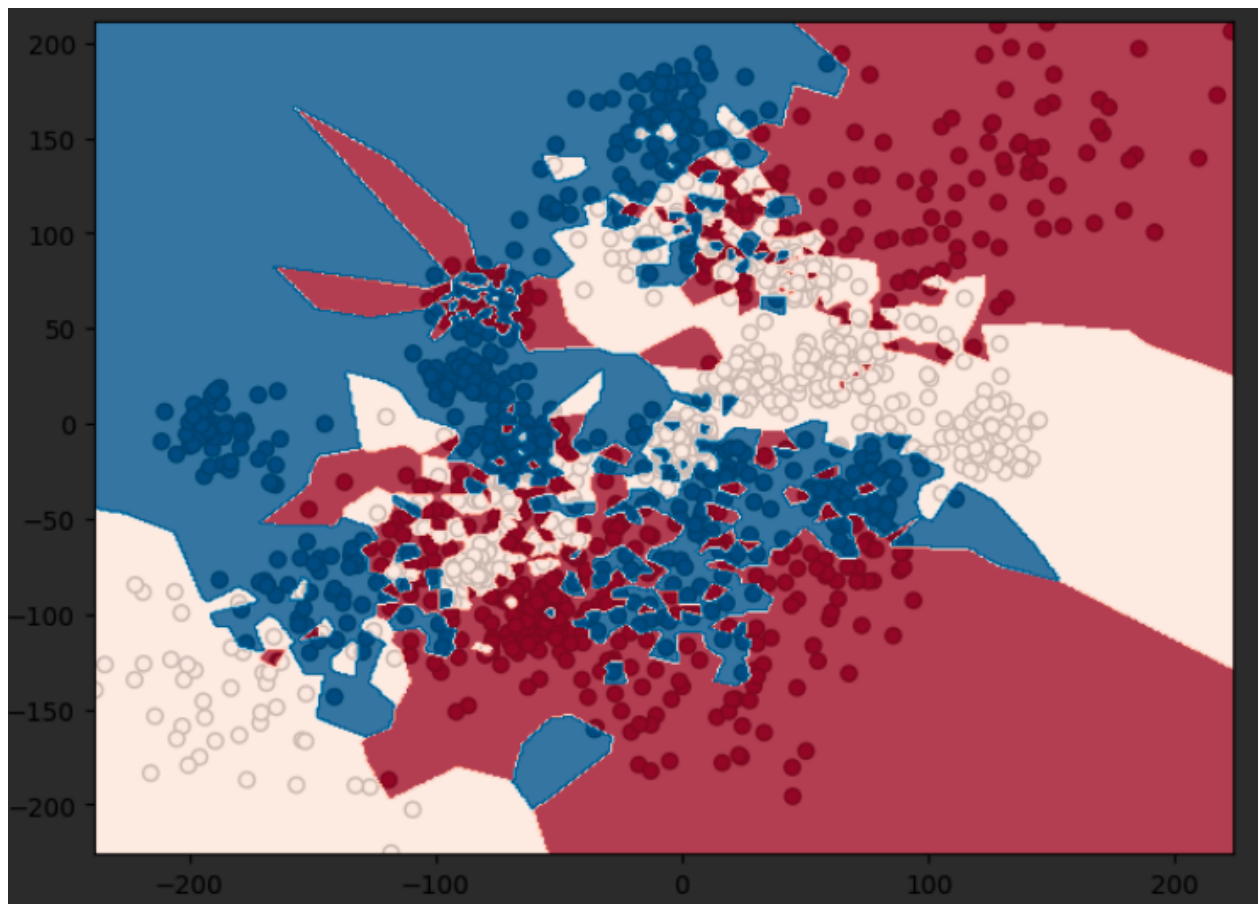
Kamil Krzempek

1 Zbiór danych

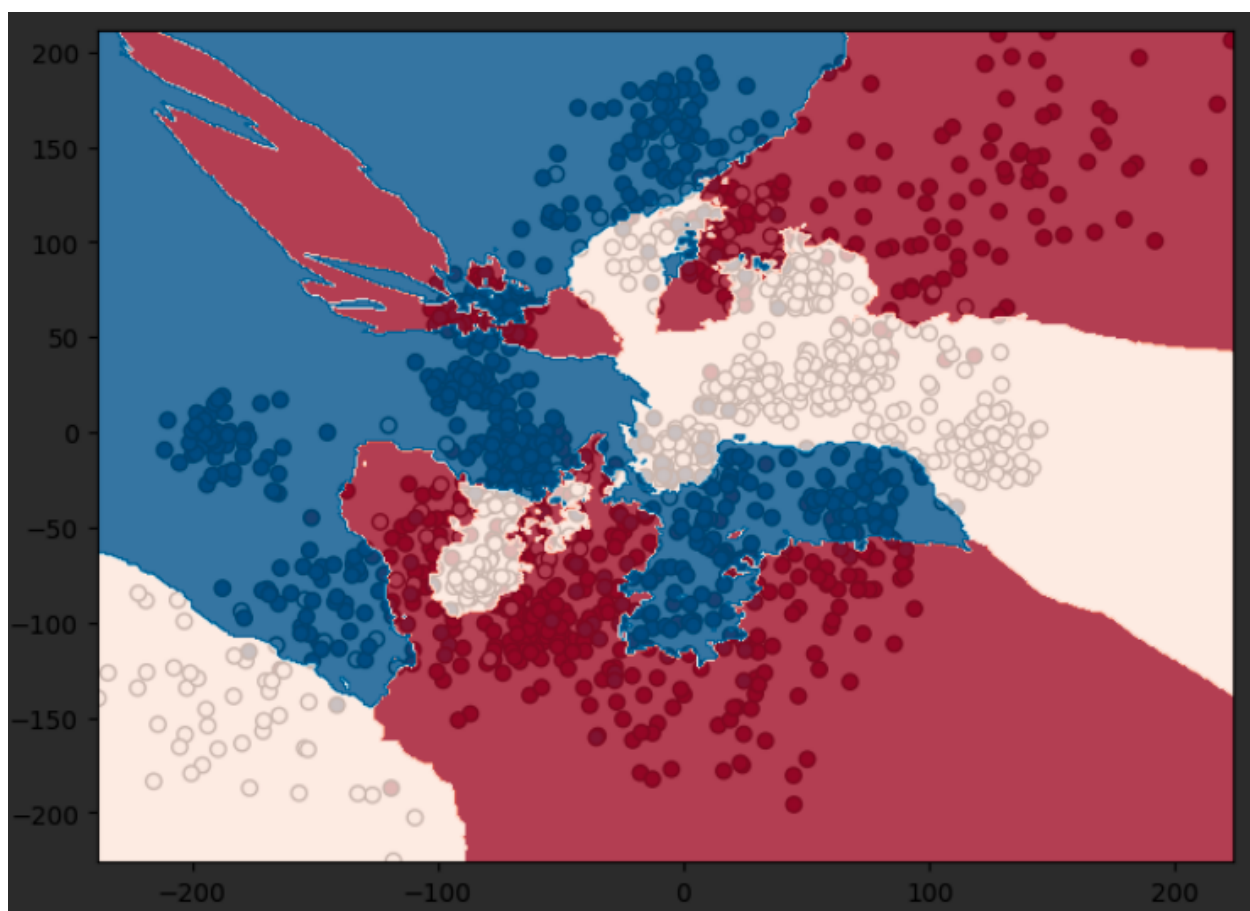


Rysunek 1: Zbiór zawiera 750 punktów i składa się z 3 klas

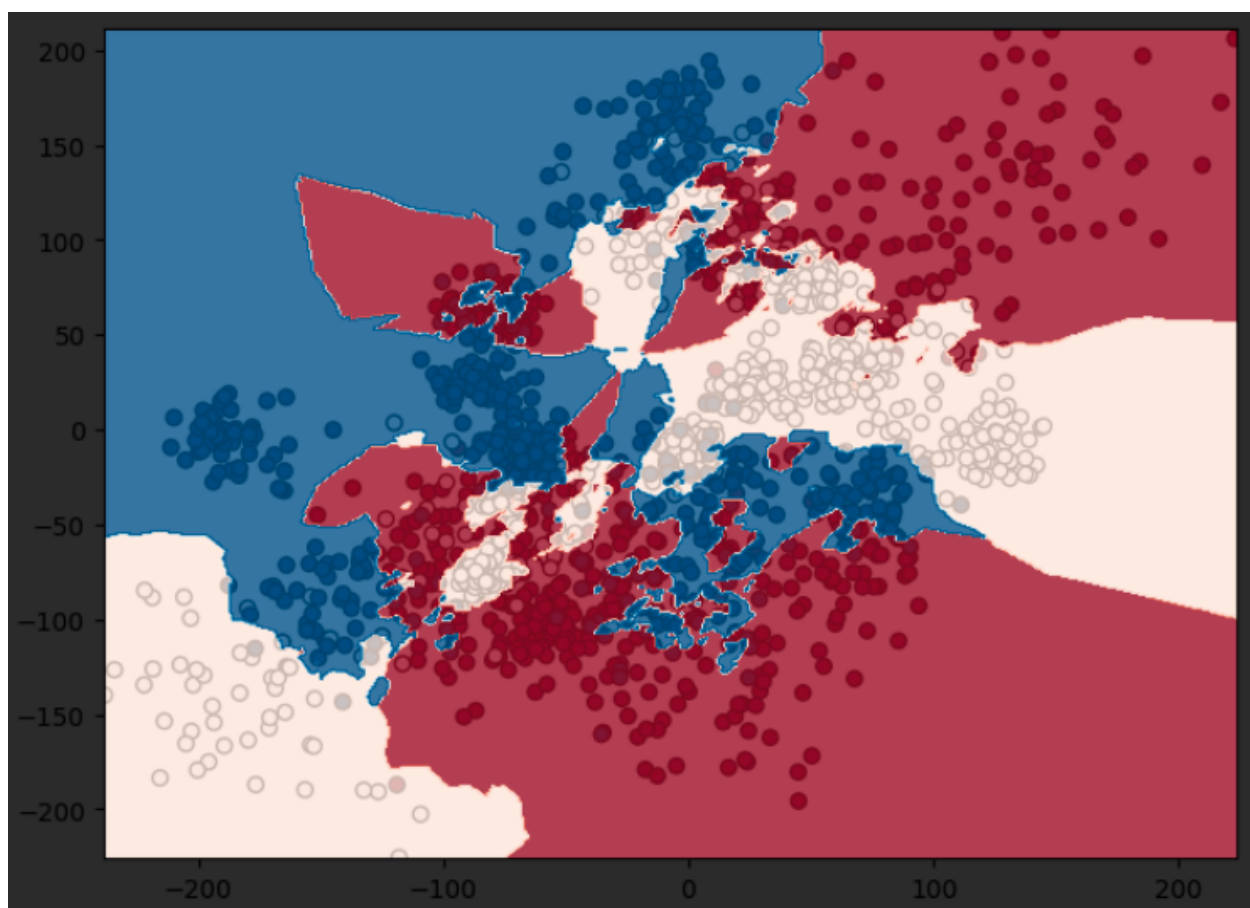
2 Granice decyzyjne



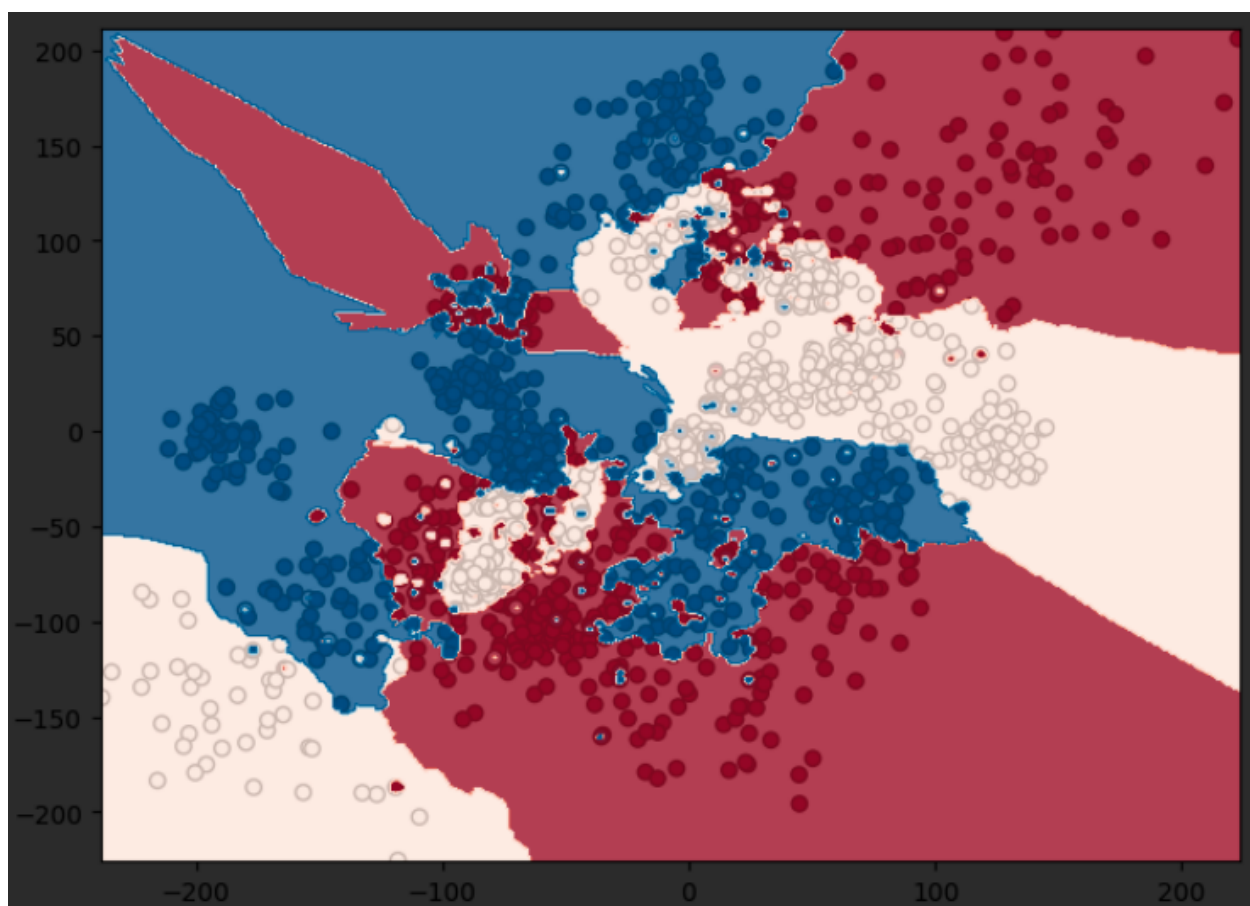
Rysunek 2: $k = 1$, metryka euklidesowa, głosowanie większościowe



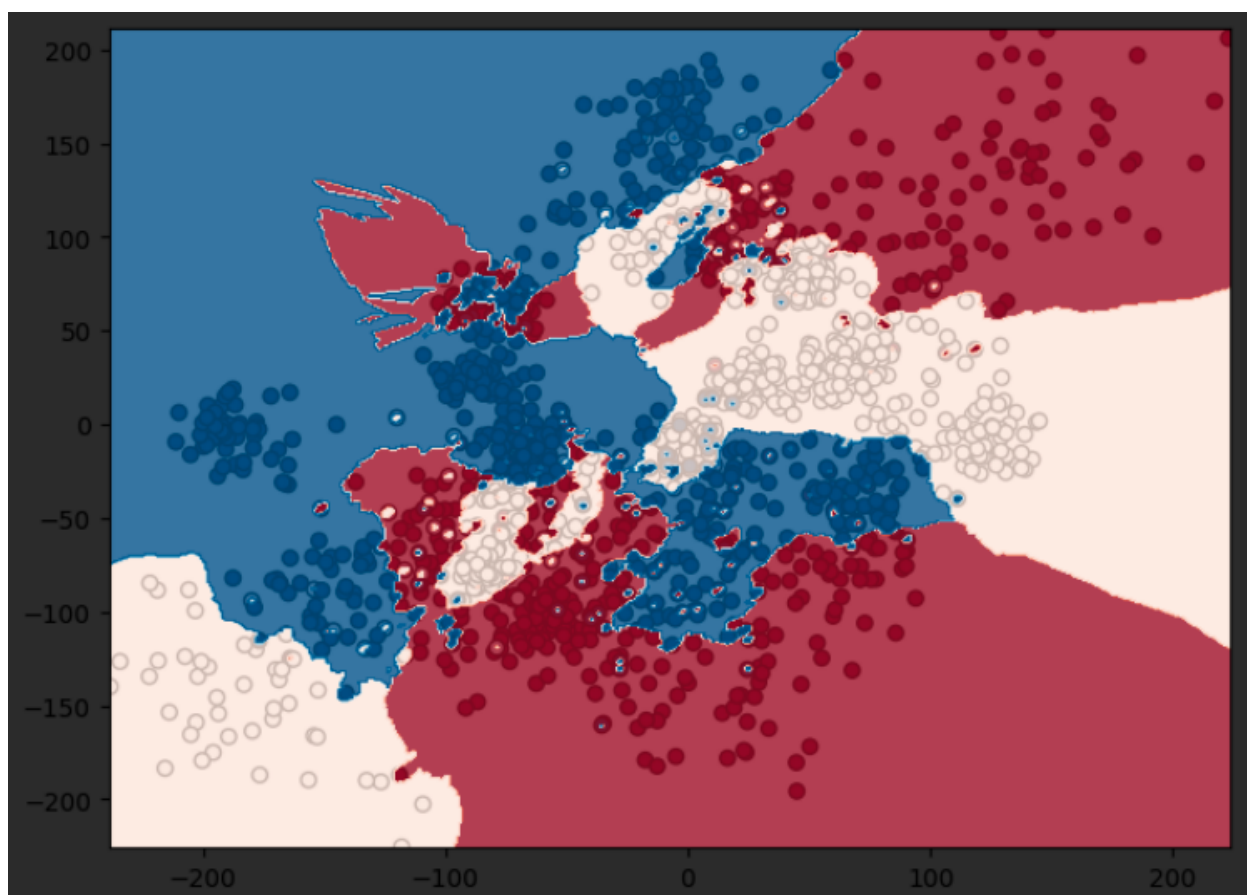
Rysunek 3: $k = 13$, metryka euklidesowa, głosowanie większościowe



Rysunek 4: $k = 4$, metryka mahalanobisa, głosowanie większościowe



Rysunek 5: $k = 9$, metryka euklidesowa, głosowanie ważone odległością



Rysunek 6: $k = 13$, metryka mahalanobisa, głosowanie ważone odległością

3 Ocena wybranych wariantów metody

Wybrano klasyfikator z głosowaniem większościowym i metryką euklidesową jako "na oko" najgorszy oraz klasyfikator z głosowaniem ważonym odległością i metryką mahalanobisa jako "na oko" najlepszy. Dla każdej wartości k uśredniono wynik po 10 próbach, najlepsze k wyznaczano po 10 razy.

Wariant z metryką euklidesową i głosowaniem większościowym uzyskał średni wynik 0.762 z odchyleniem standardowym 0.022, a drugi wariant 0.742 z odchyleniem standardowym 0.016. Są to bardzo zbliżone wartości i ciężko na ich podstawie jednoznacznie stwierdzić że jeden klasyfikator jest lepszy od drugiego.