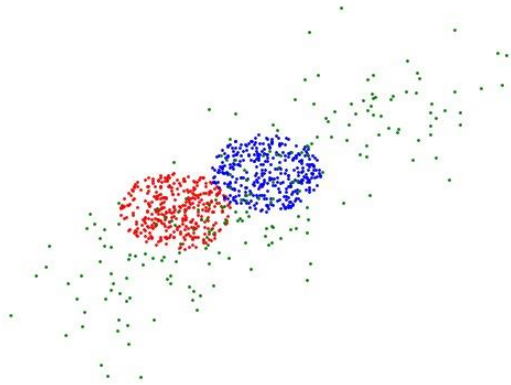
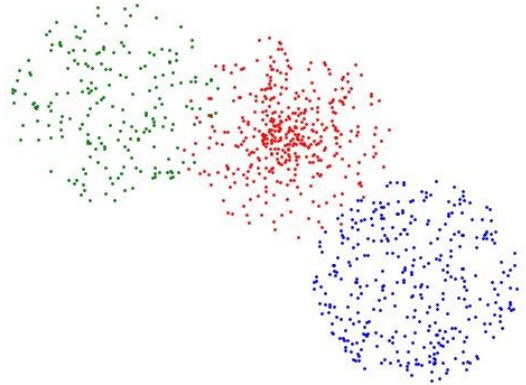


Do wykonania tego zadania stworzyłem 6 przykładowych 3 klasowych rozkładów. Zbiory były generowane przez co 2 przykładu przypadkowo jest bardzo podobny do 5.1.

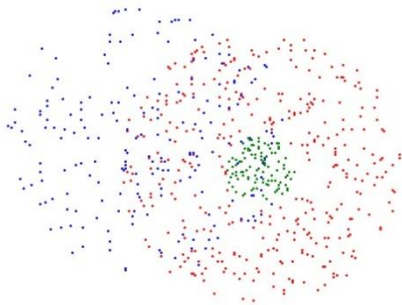
1) Nieregularny



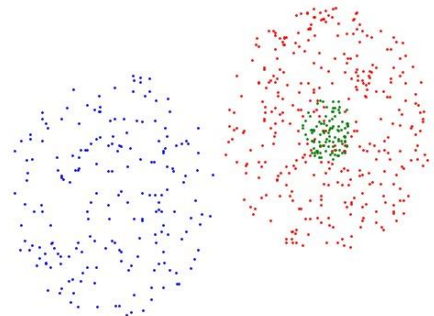
2) Różna gęstość w różnych odległościach od centrum



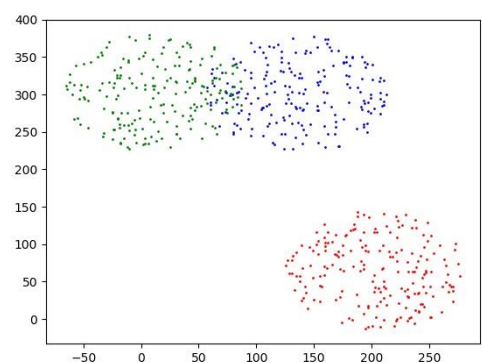
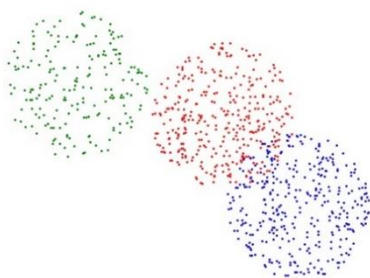
3) Jedna klasa pierścień
druga w niej schowana



4) Jedna klasa wewnątrz drugiej



5) Dwukrotnie (raz przypadkiem) jedna klasa nieznacznie nachodziła na drugą

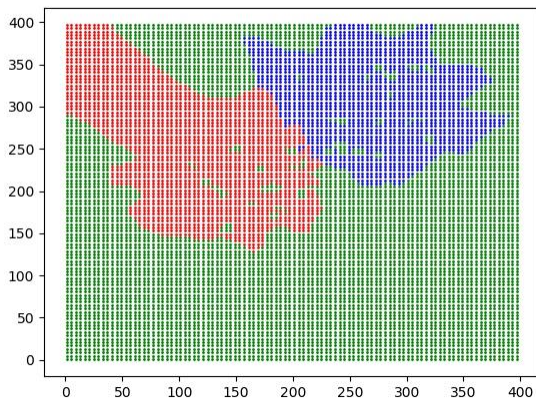


Dla każdej z sytuacji tworzyłem zalecane modele:

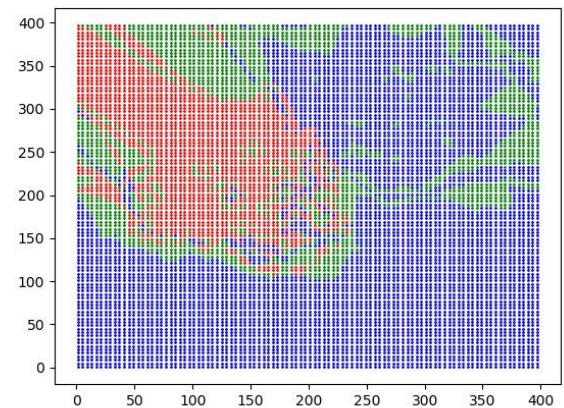
- k-NN z $k=1$, głosowaniem większościowym i metryką Euklidesa; dalej oznaczmy 1ME
- k-NN z $k=7$, głosowaniem większościowym i metryką Euklidesa; dalej oznaczmy 7ME
- k-NN z $k=7$, głosowaniem ważonym odległością i metryką Euklidesa; dalej oznaczmy 7WE
- k-NN z $k=1$, głosowaniem większościowym i metryką Mahalanobisa; dalej oznaczmy 1MM
- k-NN z $k=7$, głosowaniem ważonym odległością i metryką Mahalanobisa. dalej oznaczmy 7WM

Oto otrzymane mapy dla 1 przykładu - zawierającego nieregularny zbiór **najbardziej interesujący**

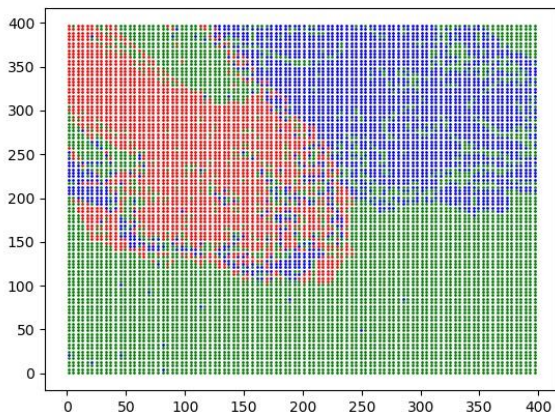
1ME (max 90%)



7ME



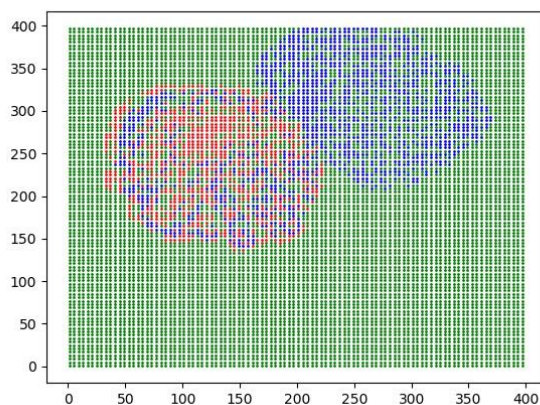
7WE



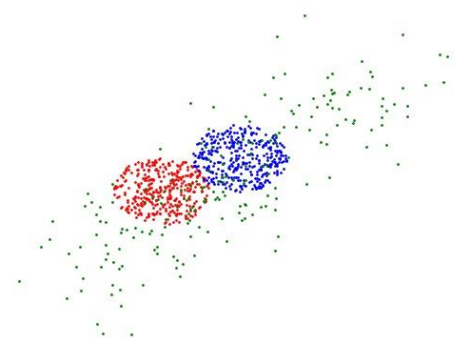
1MM



7WM (MIN 55%)

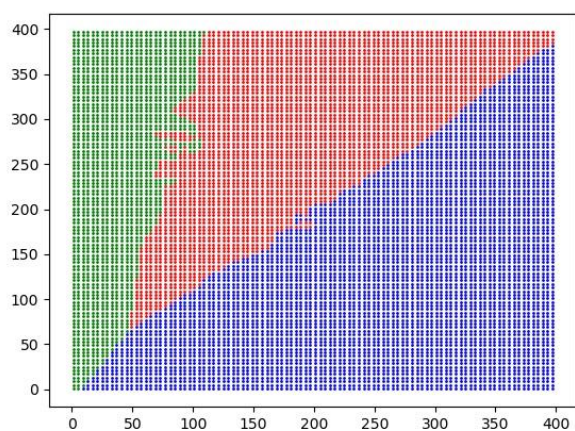


Oryginalne zdjęcie

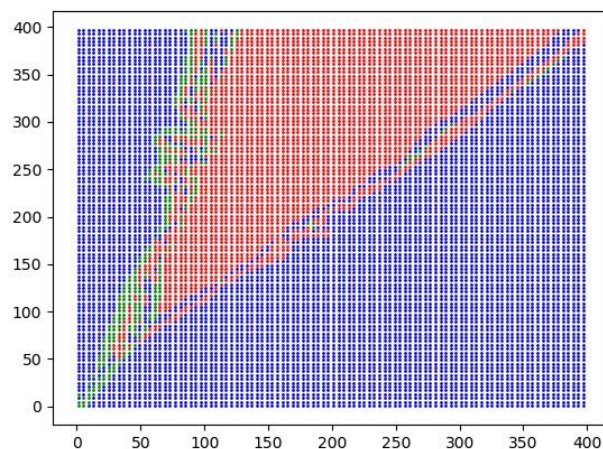


Oto otrzymane mapy dla 2 przykładu – różne gęstości

1ME



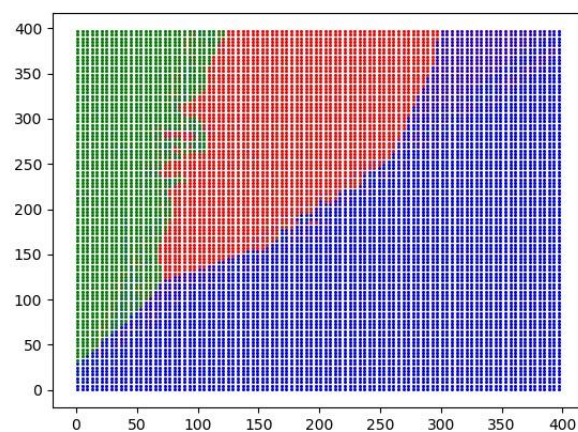
7ME (MIN 75%)



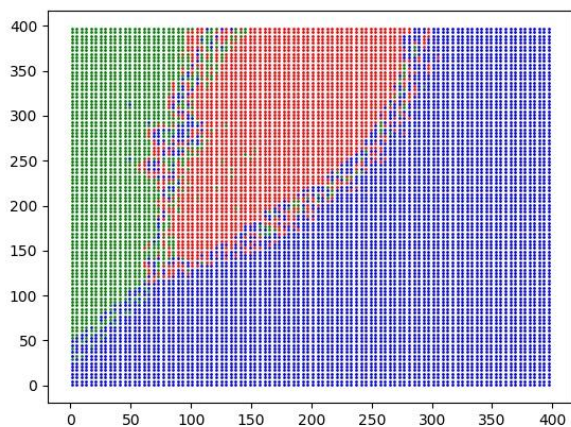
7WE



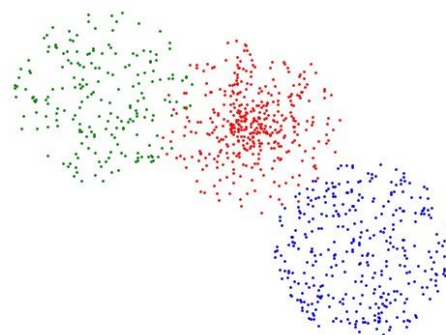
1MM (MAX 98%)



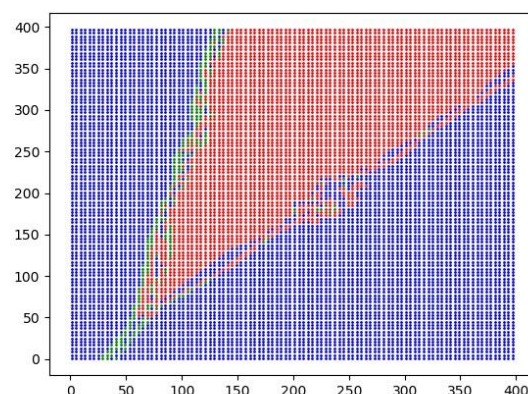
7WM



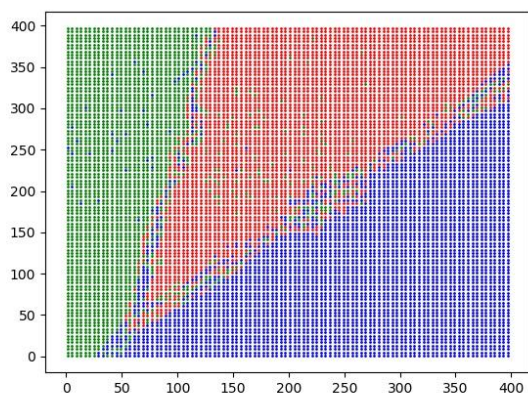
Oryginalne



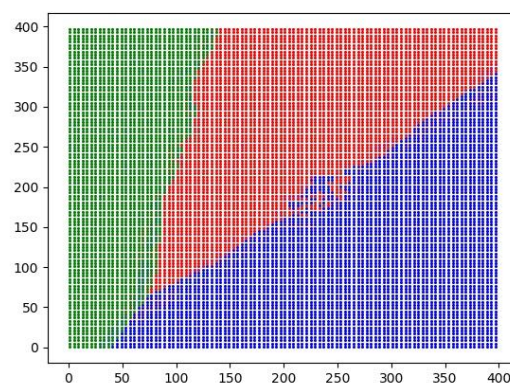
Oto otrzymane mapy dla jednego z 5 przykładów–nachodzące na siebie **drugi najbardziej interesujący**
1ME (max 95%) 7ME (min 70%)



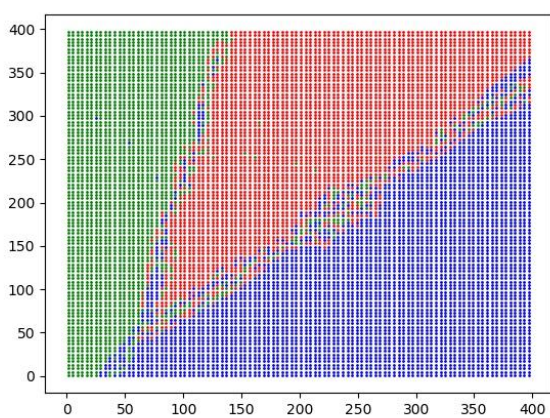
7WE



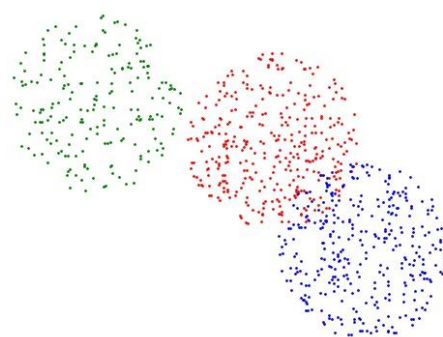
1MM



7WM



oryginalne

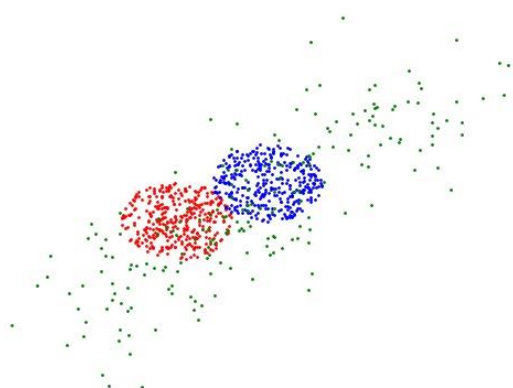


Pozostałe mapy znajdują się w przesłanej paczce, nie prezentują nic nadzwyczajnego względem powyższych.

Bardzo dziwne wydaje mi się zachowanie 7ME (7-NN most euclidean) w różnych gęstościach oraz nakładających się zbiorach, jednak jak widać na nieregularnych kształtach działa poprawnie.

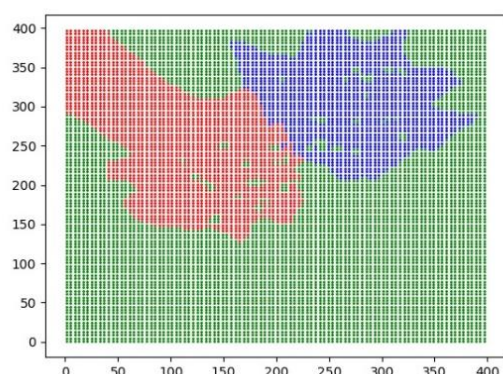
Szczegółowe wyniki dla najbardziej interesujących przypadków będą zaprezentowane dla:

- 1) Jednego nieregularnego kształtu klasy



(werble) Najlepszy wynik osiągnął (werble) : 1ME – 1NN Most Euclidean 90,05% accuracy

Macierz pomyłek dla 3 klas wygląda następująco:



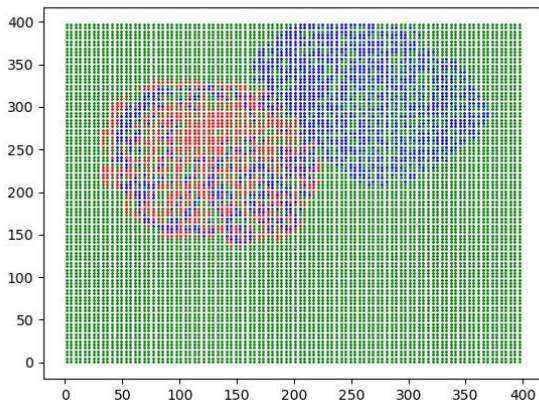
```
\      R      G      B
R      [[75.   4.   1.] #rozpoznane czerwone
G      [ 3.  27.   3.] #rozpoznane zielone
B      [ 1.   6.  61.] #rozpoznane niebieskie
```

Dla każdej klasy obliczyłem:

Precision, recall, specificity, F1 -score, Gscore

	Precision	Recall	Specifity	F1	G
R	93,75%	94,93%	95,10%	94,33%	94,34%
G	81,82%	72,97%	95,83%	77,14%	77,27%
B	89,70%	93,85%	93,97%	91,73%	91,75%

Najgorzej poradził sobie (Buuuu) : 7WM – 7-NN wage Mahalanobis 55,25% accuracy



Macierz pomyłek(confusion matrix) wygląda następująco:

```
\   R    G    B
R   [[39.  1.  1.] #rozpoznane czerwone
G   [27. 27. 30.] #rozpoznane zielone
B   [13.  9. 34.] #rozpoznane niebieskie
```

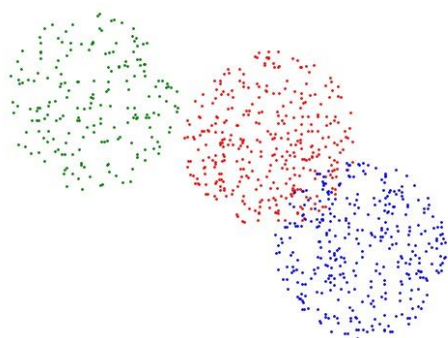
	Precision	Recall	Specifity	F1	G
R	95,12%	49,36%	98,03%	65,00%	68,53%
G	32,14%	72,97%	60,42%	44,63%	48,43%
B	60,71%	52,30%	81,03%	56,20%	56,35%

WNIOSKI:

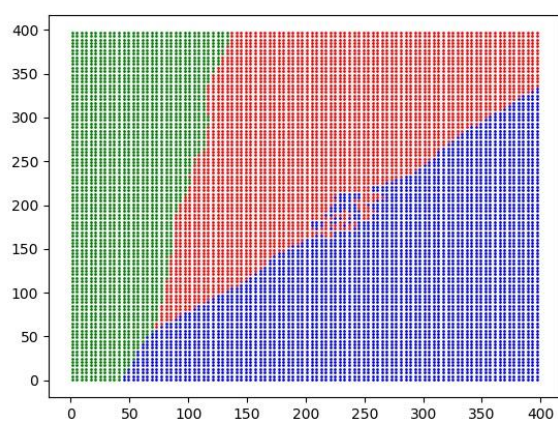
Prostszy algorytm lepiej sobie poradził w nachodzących na siebie klasach, jego skuteczność była zaskakująco wręcz wysoka. => Przy losowych zbiorach nie zawsze warto stosować macierz kowariancji jako wyznacznik odległości do danego punktu.

7MW stosunkowo dziwnie zachowało się dla 'czerwonej wyspy' na której to sklasyfikowałoby stosunkowo dużo niebieskich punktów. Czego na danych wejściowych nawet intuicyjnie nie widać.

2) Nachodzącego jednego zbioru na drugi



(werble) Najlepszy wynik osiągnął (werble) : 1ME – 1NN Most Euclidean (ponownie) 95% accuracy



Macierz pomyłek(confusion matrix) wygląda następująco:

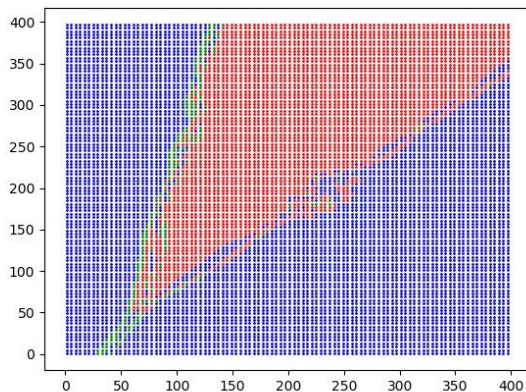
```

\   R   G   B
R   [[66.  0.  4.] rozpoznane czerwone
G   [ 0. 46.  0.] rozpoznane zielone
B   [ 7.  0. 79.] rozpoznane niebiesk

```

	Precision	Recall	Specifity	F1	G
R	94,29%	90,41%	98,90%	92,31%	92,33%
G	100%	100%	100%	100%	100%
B	91,86%	95,18%	94,12%	93,49%	93,51%

Najgorzej poradził sobie: 7ME - 7-NN Most euclidean 70% accuracy



Macierz pomyłek(confusion matrix) wygląda następująco:

```
\ R   G   B
R [[62.  1.  4.] Rozpoznane czerwone
G [  2.  0.  0.] Rozpoznane zielone
B [  9. 45. 79.] Rozpoznane niebieskie
```

	Precision	Recall	Specifity	F1	G
R	92,54%	84,93%	96,12%	88,57%	88,65%
G	0%	0%	98,72%	0 DIVISION	0%
B	59,40%	95,18%	54,62%	73,15%	75,19%

WNIOSKI/PRZEMYŚLENIA:

Ponownie najprostszy algorytm wykazał najlepszą skuteczność. Możliwe, że wygenerowane zbiory faworyzują ten typ klasyfikacji

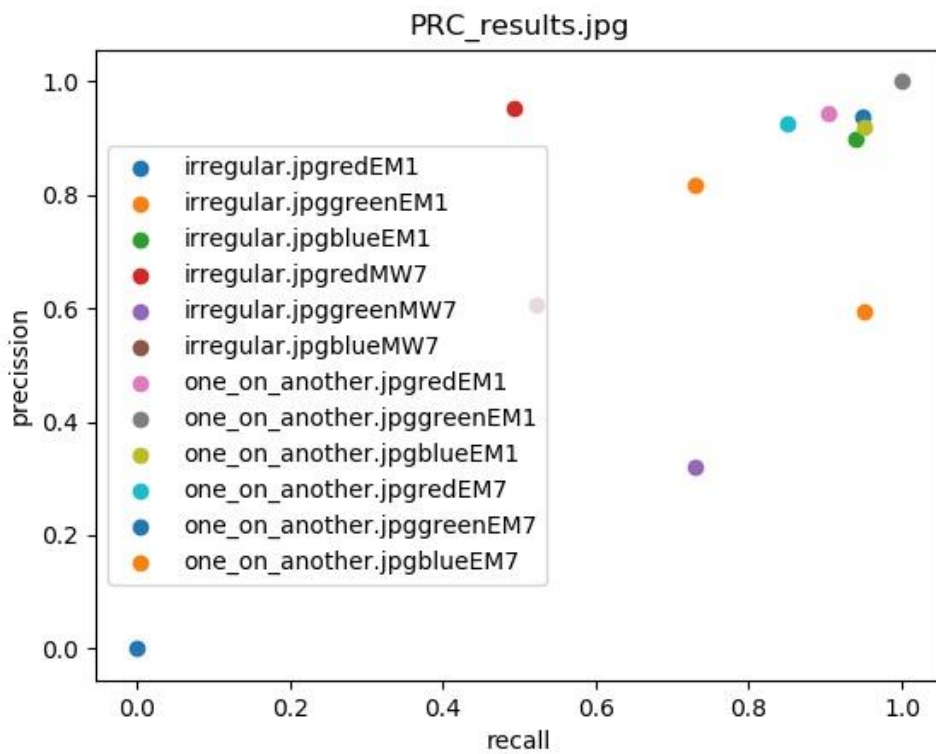
7ME zachował się dla mnie bardzo zaskakująco w lewej części mapy analizowałem wielokrotnie to zjawisko szukając swojego błędu jednak go nie odkryłem.

W7E widać, że największy problem sprawiła klasyfikacja zielonego koloru. (przy założeniu poprawności działania).

Co ciekawe dla bardzo podobnego spektrum danych (różnej gęstości), prezentowanej wcześniej, najlepiej poradził sobie MM1 – 1-NN most Mahalanobis osiągając accuracy w wysokości 98,4%

Najważniejszym wnioskiem z całego eksperymentu według mnie jest to, że dla różnych zbiorów warto używać różnych klasyfikatorów. Czasem warto sprawdzić kilka aby wybrać ten właściwy.

Na bazie powyższych wyników sporządziłem wykres PRC oraz ROC. Sporządziłem je rozróżniając klasyfikację w zależności od koloru. (przepraszam za brzydką legendę).



Sporządzone wykresy pozwalają na łatwiejszą i szybszą interpretację wyników przedstawionych powyżej. Są one wprost od nich zależne.

