

Gliwice, 2025-06-08



Politechnika Śląska

Wydział Automatyki, Elektroniki i Informatyki
Informatyka

Wizja komputerowa i rozpoznawanie obrazów
Narzędzia do ewaluacji klasyfikatorów

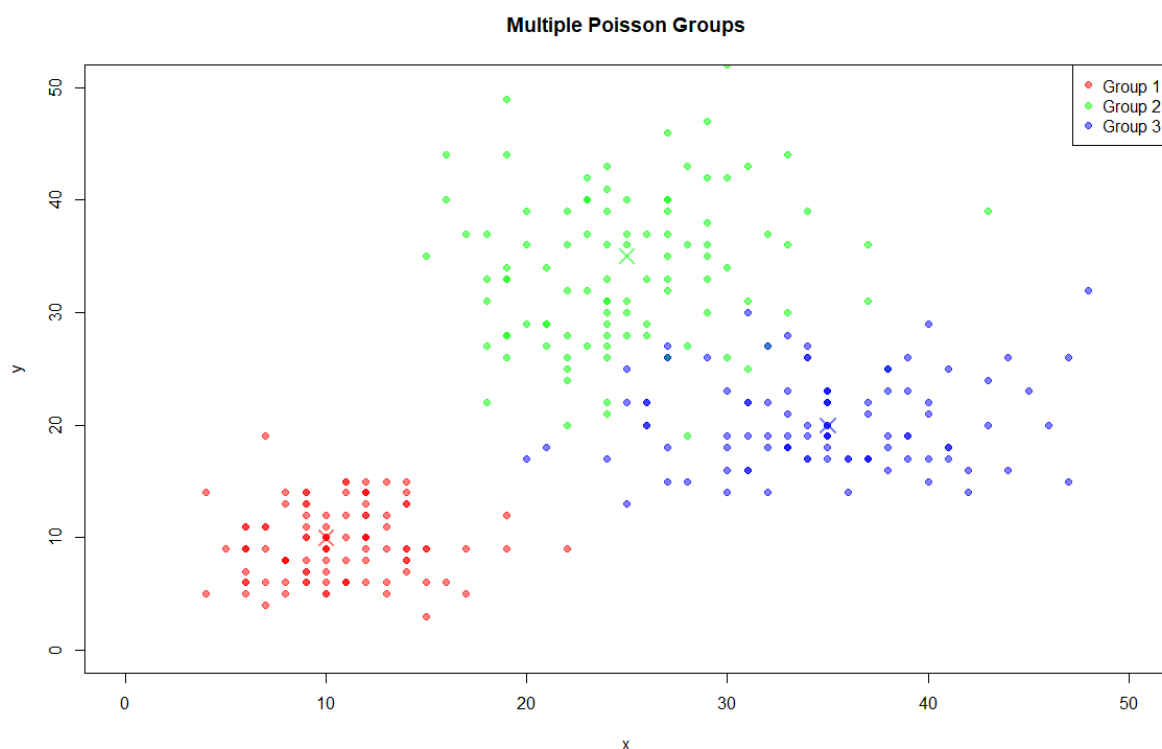
Wojciech Adamczyk
Tomasz Artmański
Dominik Brzezina
Karol Galański
Piotr Kraska
Bartłomiej Mikoda

Wprowadzenie

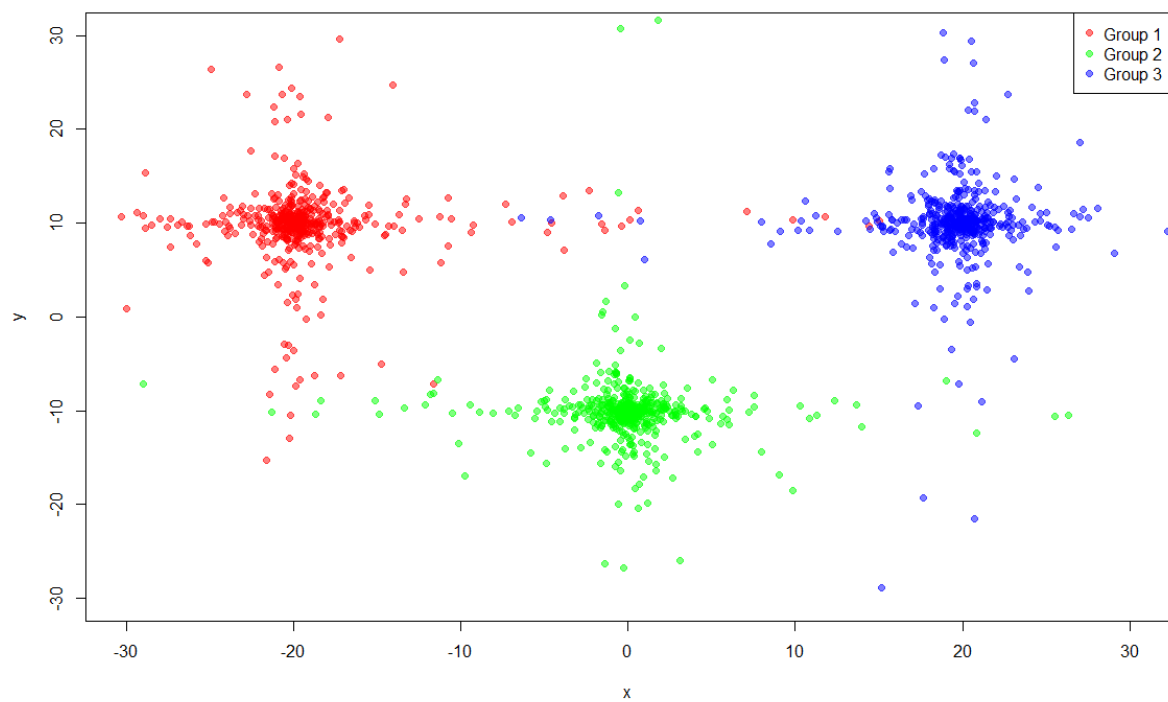
W ramach projektu mieliśmy za zadanie zapoznać się z artykułem zamieszczonym w publikacji *Applied Soft Computing Journal*, zatytułowanym „*Classification Assessment Tool: A program to measure the uncertainty of classification models in terms of class-level metrics*”. Następnie mieliśmy oprzeć się na ustaleniach autorów artykułu i przeprowadzić eksperymenty ustalające dokładność poszczególnych klasyfikatorów na testowych zbiorach danych. Podsumowaniem projektu było sporządzenie sprawozdania podsumowującego przebieg prac nad projektem oraz nasze ustalenia.

Generacja zbiorów danych

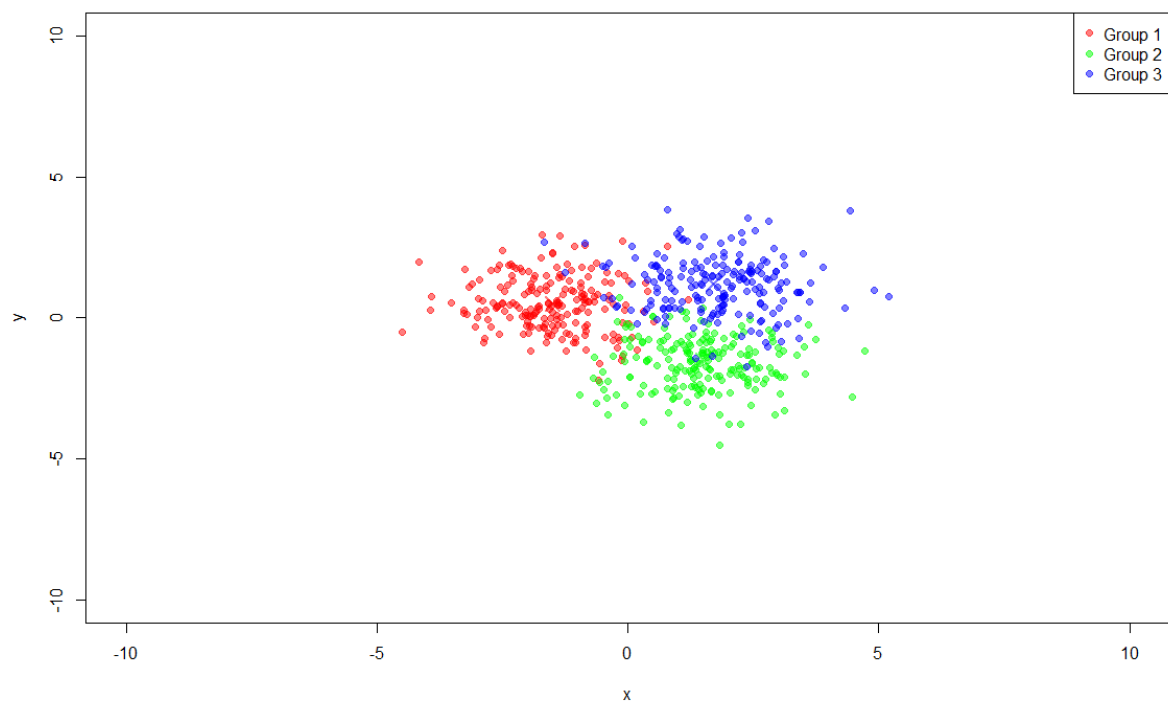
Pierwszym krokiem w projekcie było stworzenie algorytmów pozwalających nam na generację zbiorów danych, na których moglibyśmy przeprowadzić dalsze eksperymenty. Napisaliśmy w tym celu funkcje umożliwiające nam tworzenie testowych zbiorów danych w dystrybucji Poisson’a, Cauchy’ego, oraz Gauss’a. Poniżej kilka przykładowych zbiorów:



Multiple Cauchy Groups

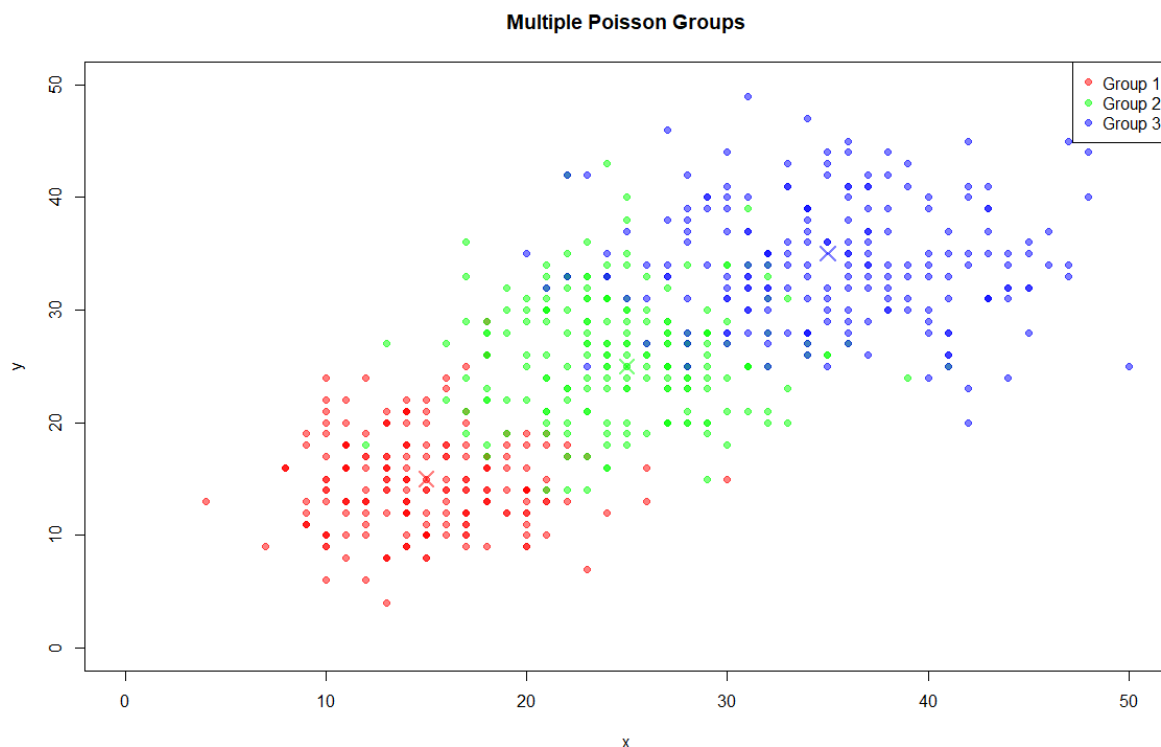


Multiple Gauss Groups



Przykładowe działania na zbiorach danych, objaśnienie

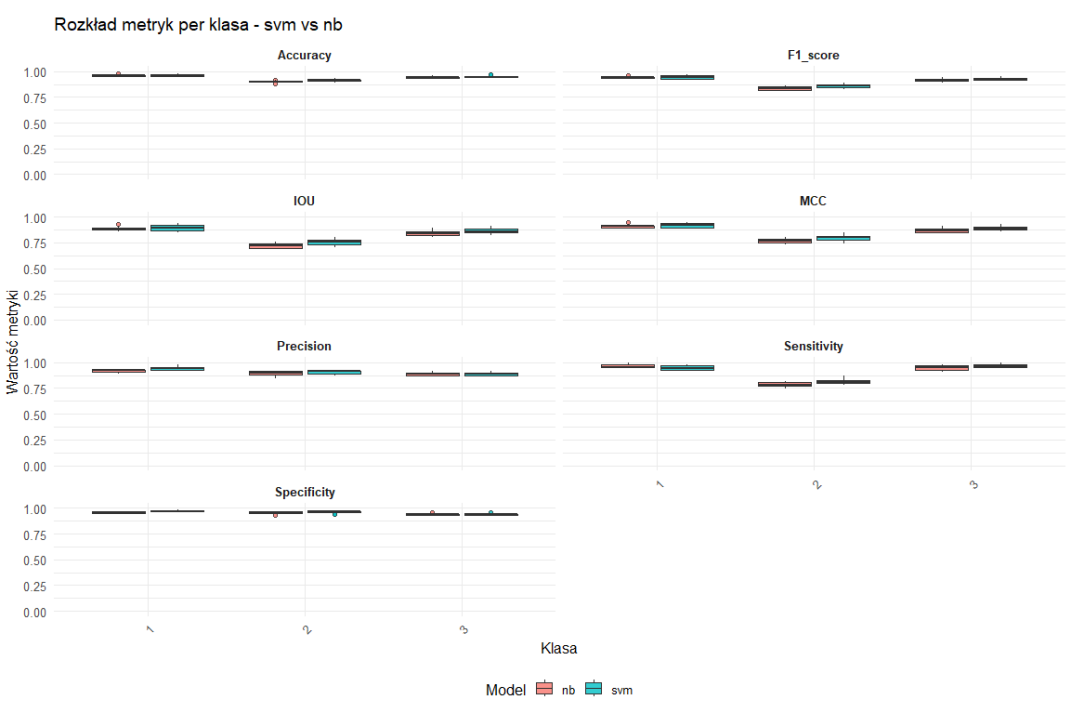
Na przykładzie pierwszego zestawu danych przedstawimy pokrótce operacje wykonywane w celu uzyskania danych na temat skuteczności poszczególnych klasyfikatorów. Oto nasz zestaw danych:



W pierwszej kolejności stosujemy na nim klasyfikator SVM oraz naiwny klasyfikator Bayes'a. Następnie określamy metryki dla każdej możliwej kombinacji grupa-klasyfikator. Do badanych metryk należą:

1. Accuracy - Procent poprawnych przewidywań względem wszystkich przykładów.
2. Sensitivity (aka. Recall, True Positive Rate) - Odsetek rzeczywistych pozytywnych przypadków, które zostały poprawnie wykryte.
3. Specificity (True Negative Rate) - Odsetek rzeczywistych negatywnych przypadków, które zostały poprawnie wykryte.
4. F1 Score - Średnia harmoniczna precision i sensitivity.
5. IoU (Intersection over Union) - Miara nakładania się obszarów.
6. MCC (Matthews Correlation Coefficient) - Korelacja między przewidywaniami a rzeczywistymi klasami.
7. Precision - Odsetek poprawnych pozytywnych przewidywań wśród wszystkich przewidzianych jako pozytywne.

Dla naszego przykładowego zbioru danych otrzymaliśmy następujące metryki:



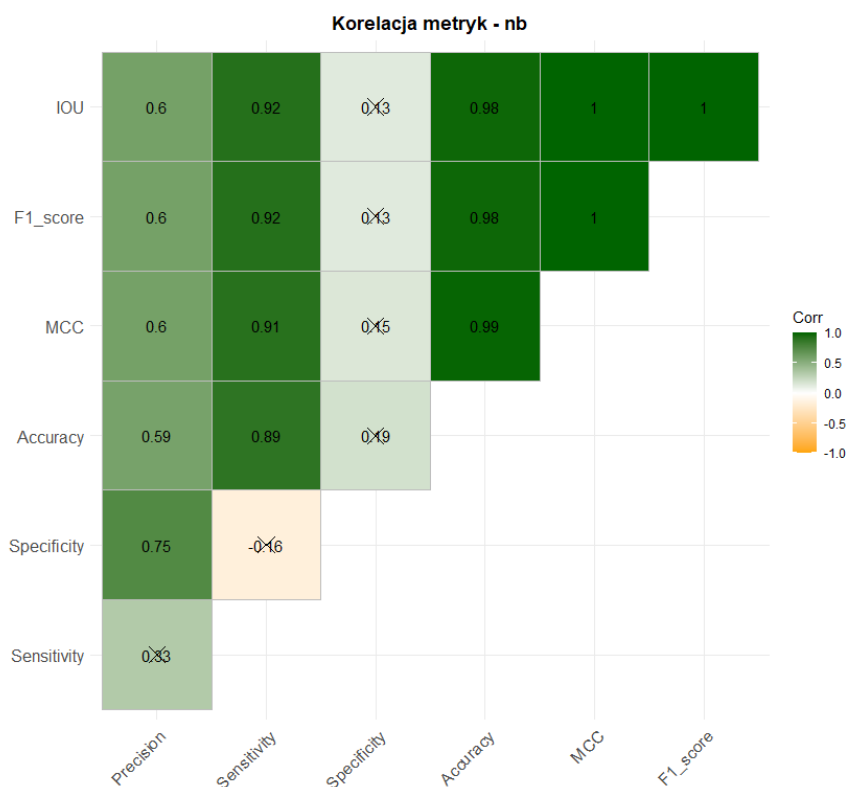
Każda metryka zawiera informację o rozkładzie otrzymanych wartości dla danej grupy, z wyróżnieniem poszczególnych klasyfikatorów.

Następnym krokiem jest generacja mapy cieplnej ilustrującej nasze wyniki dla różnych parametrów testowych. Poniżej przykładowa heatmapa metryki Accuracy dla naiwnego klasyfikatora Bayes'a:



Oś X reprezentuje tutaj ułamek pełnego zestawu danych, jaki jest brany pod uwagę przy każdym teście (przykładowo Fraction 0.6 oznacza, że losowo wybieramy 60% punktów danych). Miarą osi Y jest ilość testów, jakie wykonujemy – teoretycznie im więcej testów, tym większa precyzja, choć nie sprawdza się to zawsze.

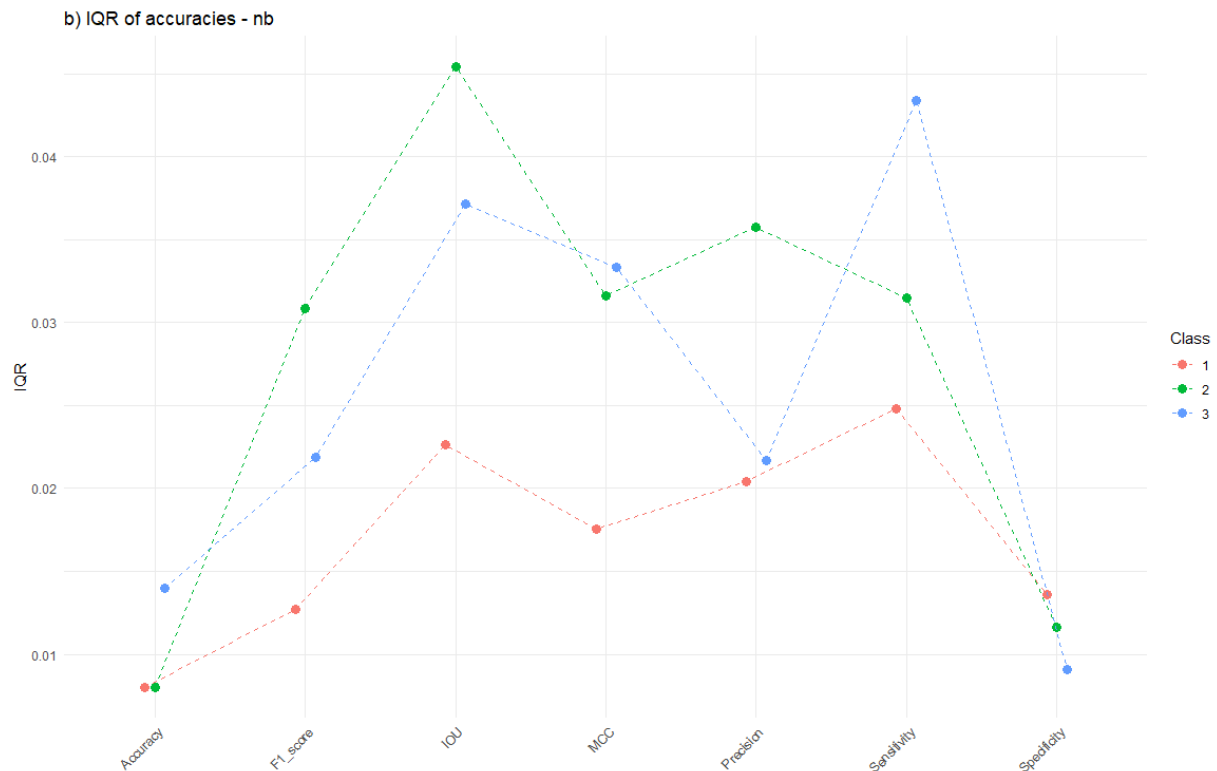
W kolejnym kroku tworzymy wykres ilustrujący korelację metryk, jaka występuje w naszym zestawie danych:



Wraz z oddaleniem danej wartości korelacji od zera wzrasta wpływ, jaki wzajemnie wywierają na siebie konkretne metryki. Przykładowo, korelacja pomiędzy IOU oraz Accuracy jest wysoka oraz dodatnia, co oznacza, że metryki te mają na siebie bardzo mocny wpływ – ze wzrostem jednej z nich obserwujemy porównywalny wzrost drugiej. Z kolei korelacja między Specificity oraz Sensitivity jest niska i ujemna – ze wzrostem jednej z nich obserwujemy nieznaczny spadek drugiej.

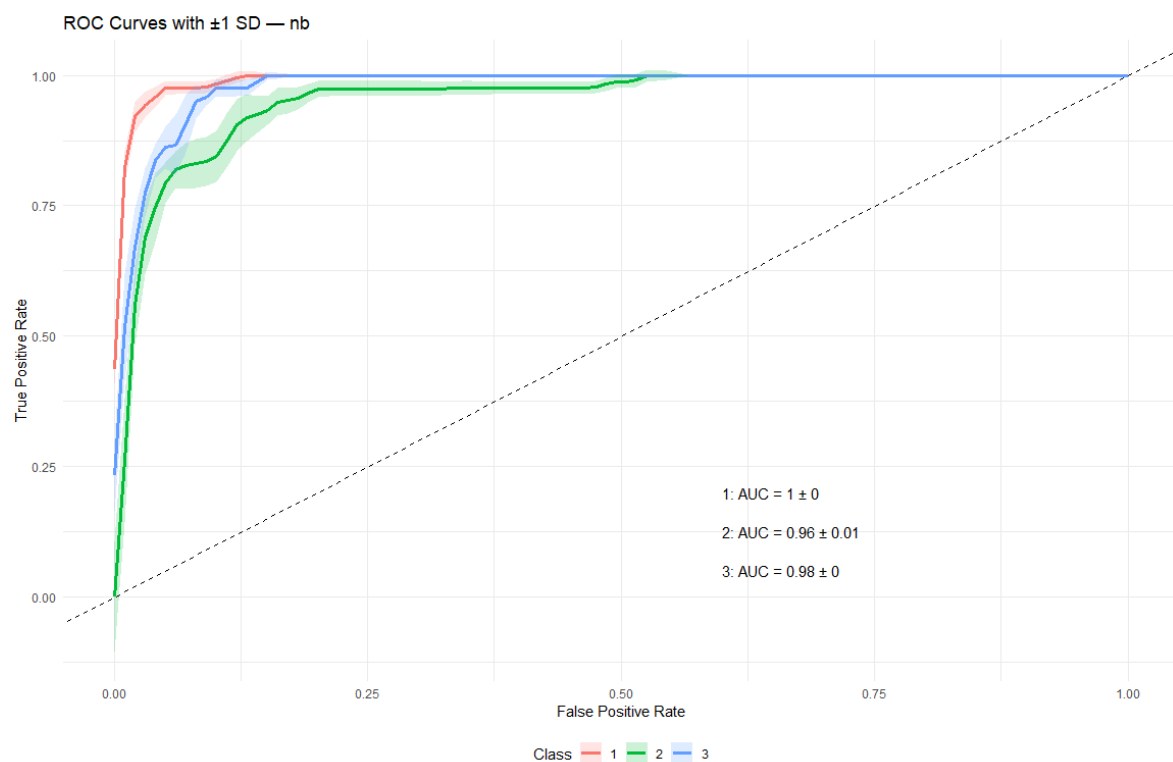
Kolejnym krokiem w przetwarzaniu naszych danych jest wygenerowanie wykresu IQR (Interquartile Range). Jest to różnica między górnym kwartylem (75. percentyl) a dolnym kwartylem (25. percentyl) danych, w naszym wypadku poszczególnych metryk. IQR oblicza się w celu określenia rozrzutu środkowych 50% danych.

Na kolejnej stronie widoczny jest wykres IQR dla naiwnego klasyfikatora Bayes'a ,z podziałem pomiędzy poszczególne klasyfikatory oraz grupy w zbiorze danych.



Ostatnią operacją jest generacja wykresu ROC (Receiver Operating Characteristic) – wartość ta służy do oceny skuteczności klasyfikatora binarnego – szczególnie takiego, który zwraca prawdopodobieństwo przynależności do klasy pozytywnej. Wykres ROC ilustruje zależność pomiędzy True Positive Rate a False Positive Rate.

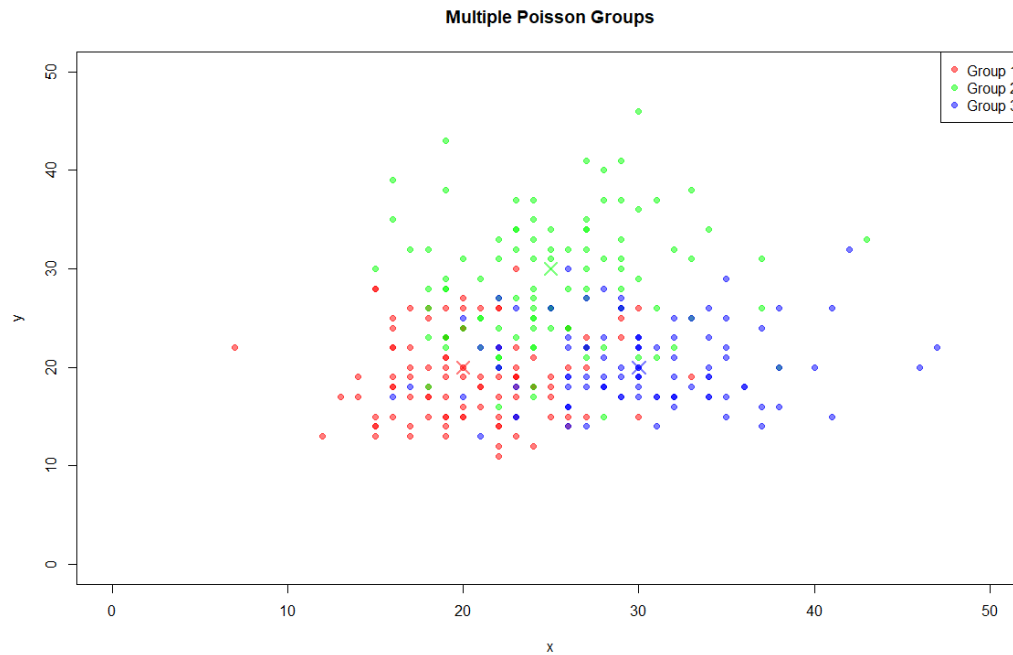
Oto przykładowy wykres ROC dla danych, z którymi dotychczas działaliśmy:



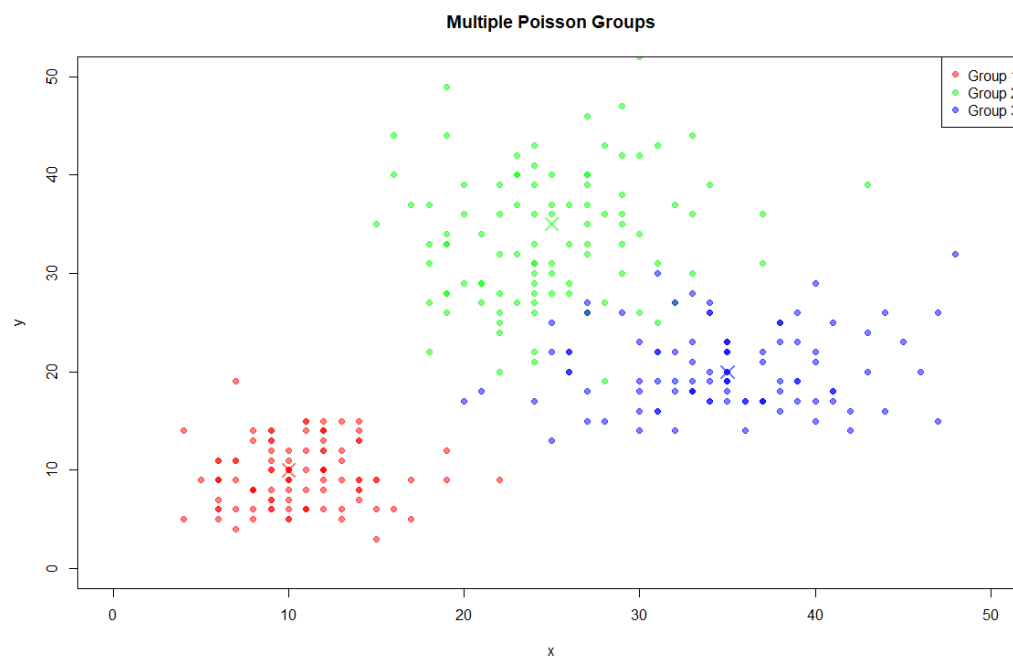
Dane testowe

Poniżej widoczne jest sześć zestawów danych, na których zostanie przez nas przeprowadzona ewaluacja klasyfikatorów. Dwa z nich wykorzystują dystrybucję Poisson'a, kolejne dwa dystrybucję Cauchy'ego, a ostatnie dwa dystrybucję Gauss'a.

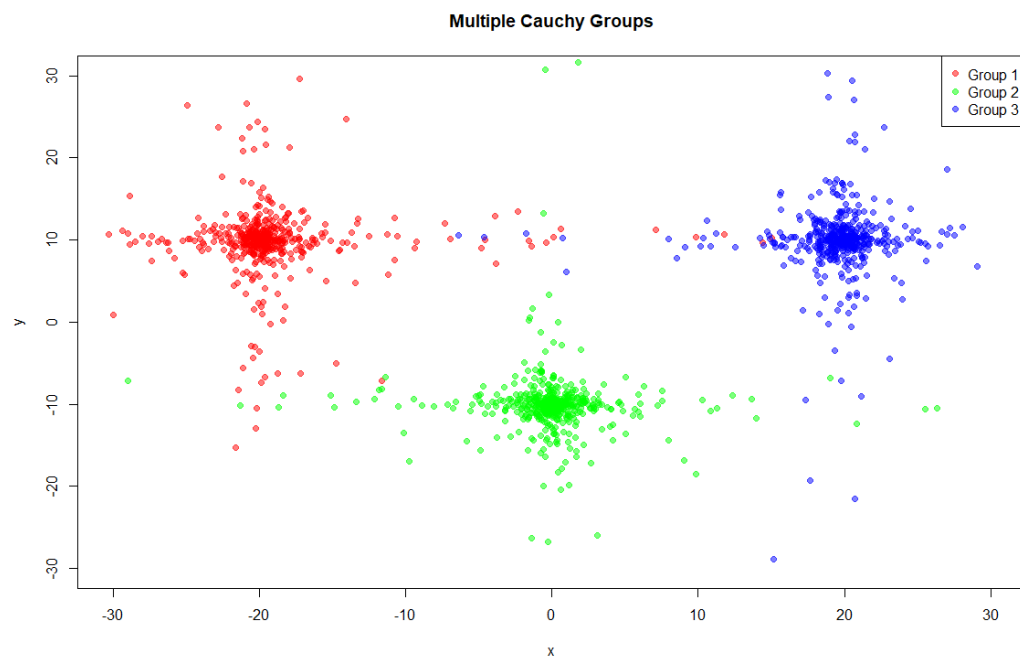
Zestaw 1:



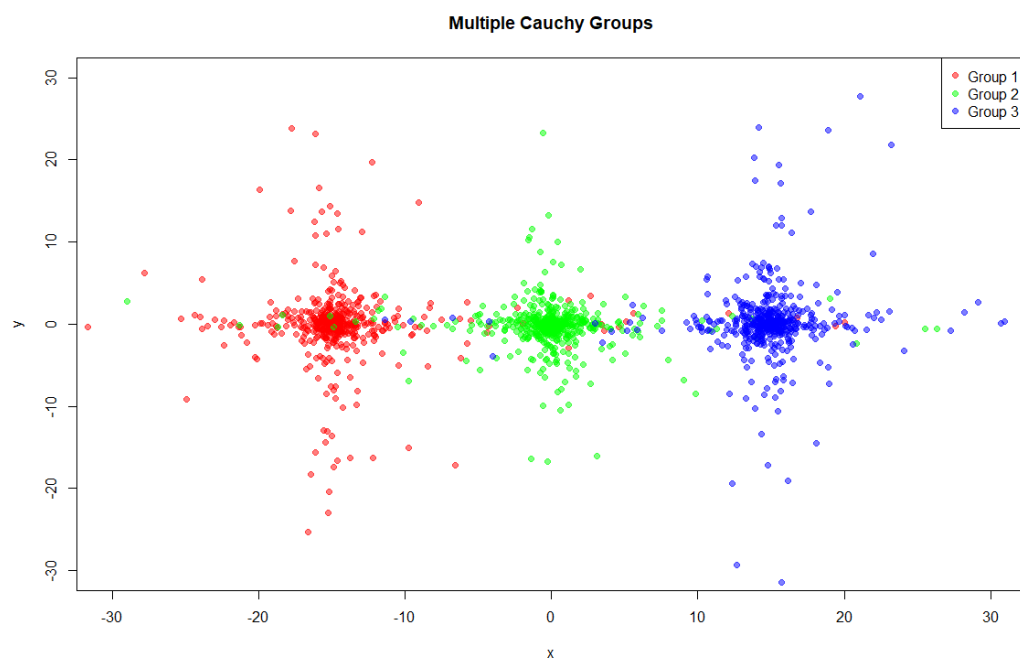
Zestaw 2:



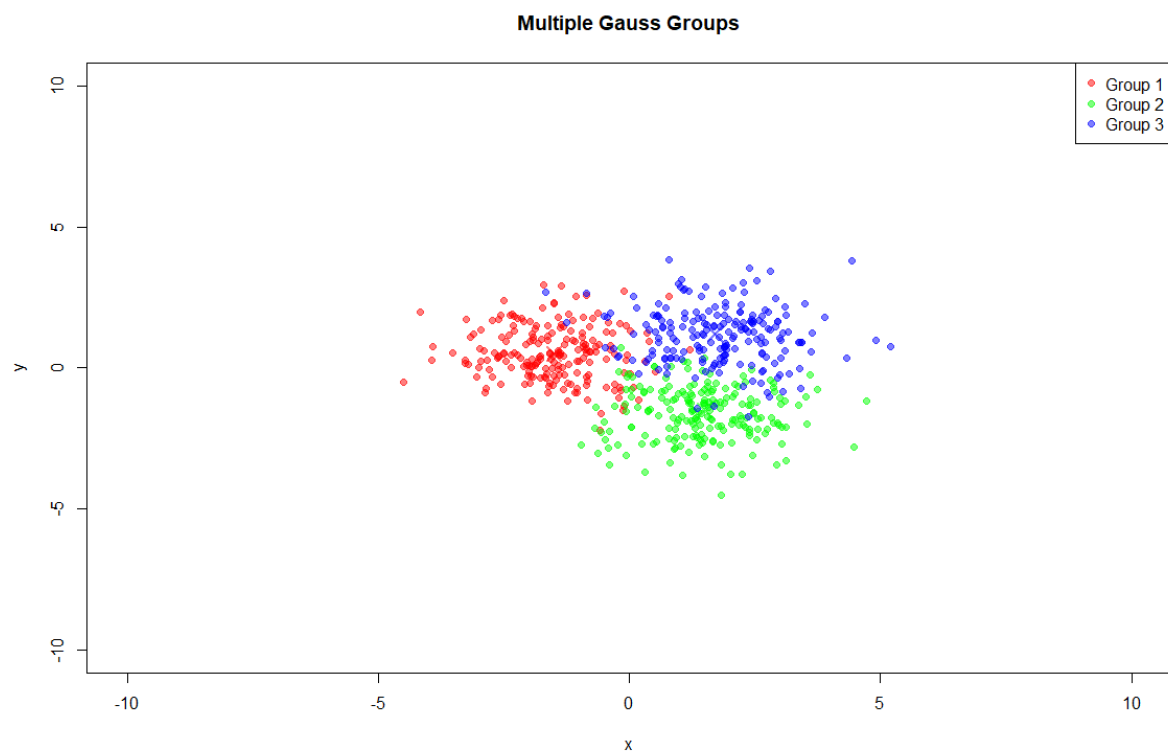
Zestaw 3:



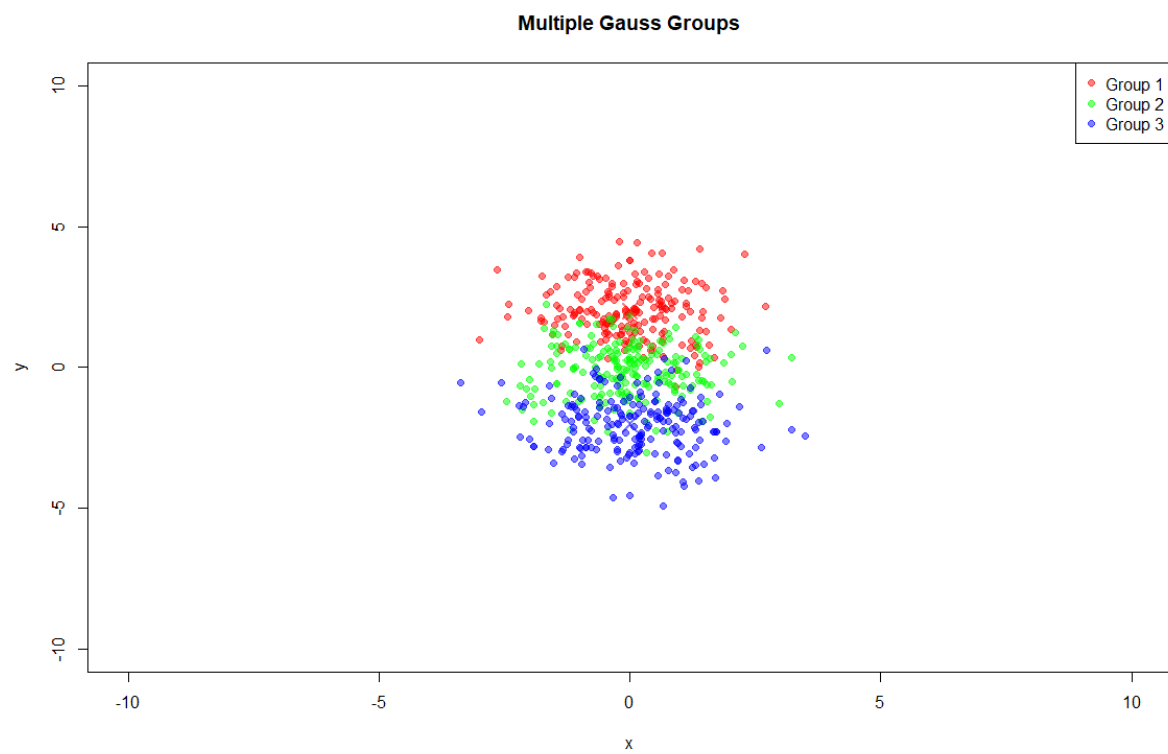
Zestaw 4:



Zestaw 5:

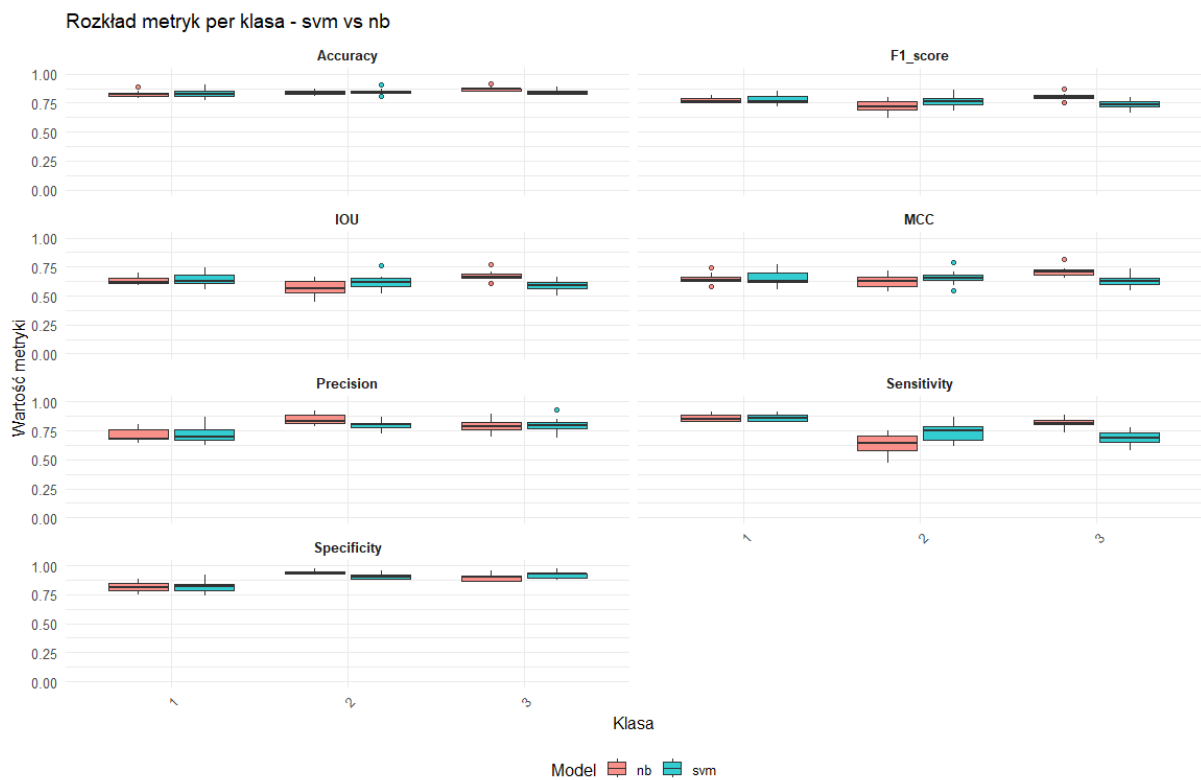
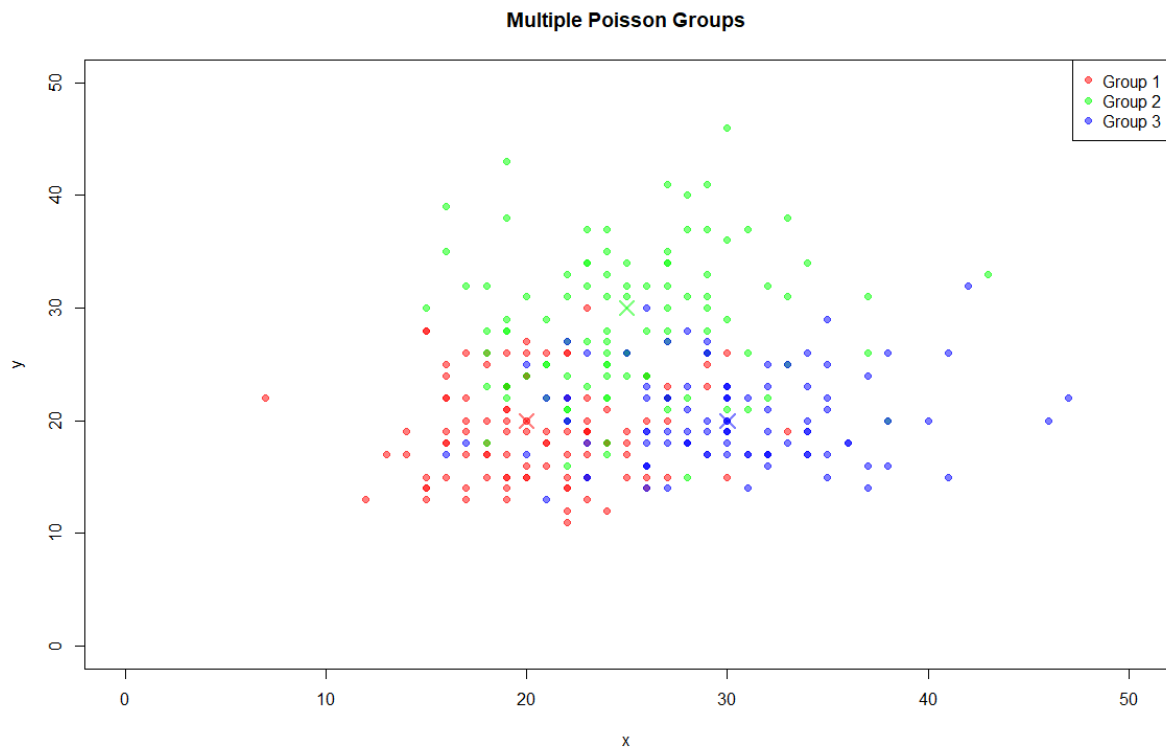


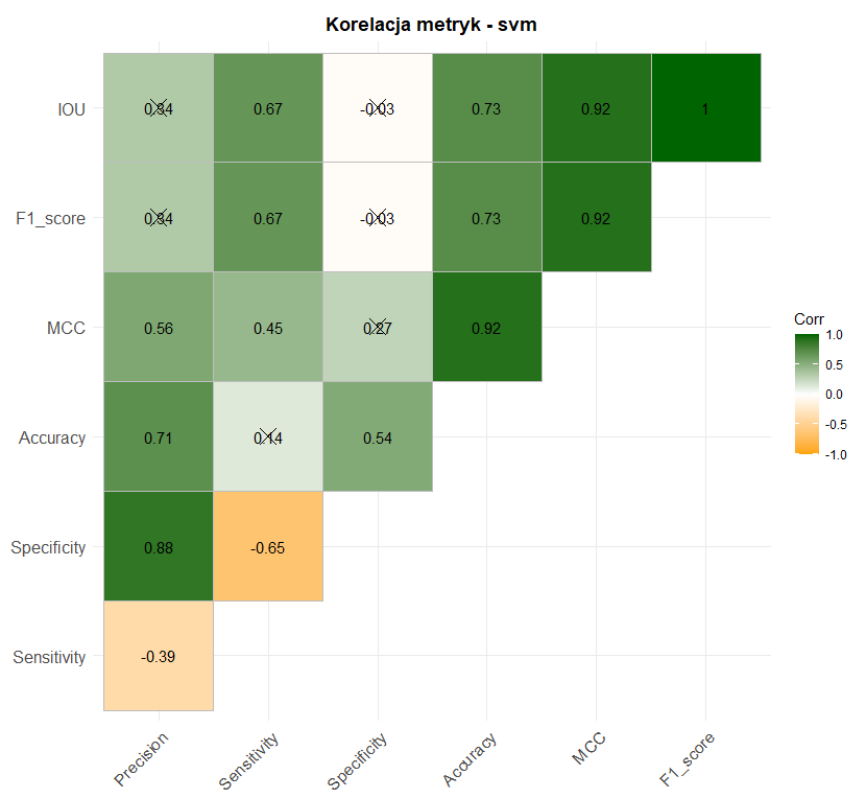
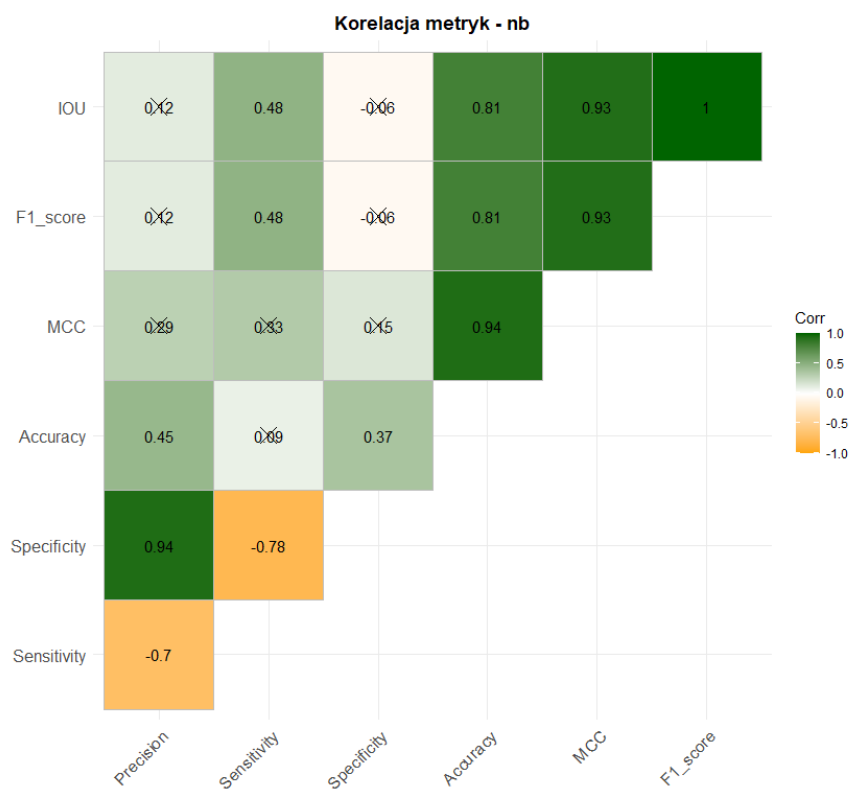
Zestaw 6:

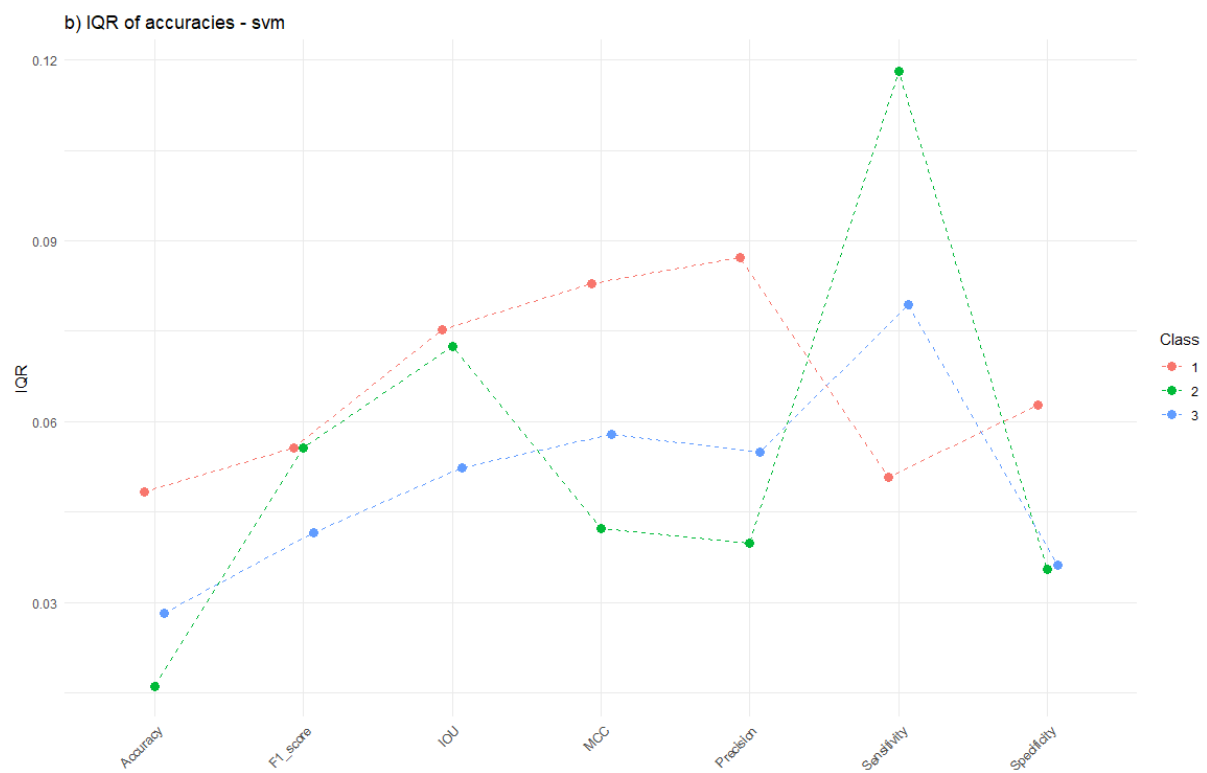
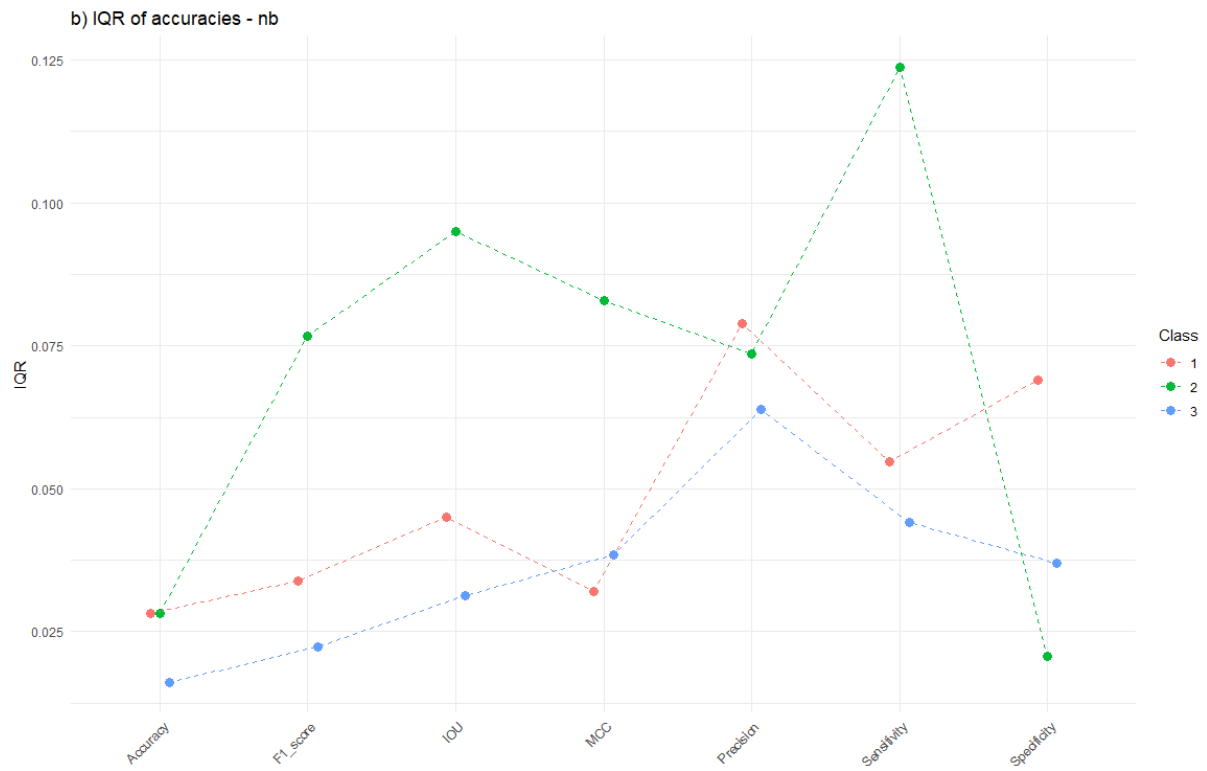


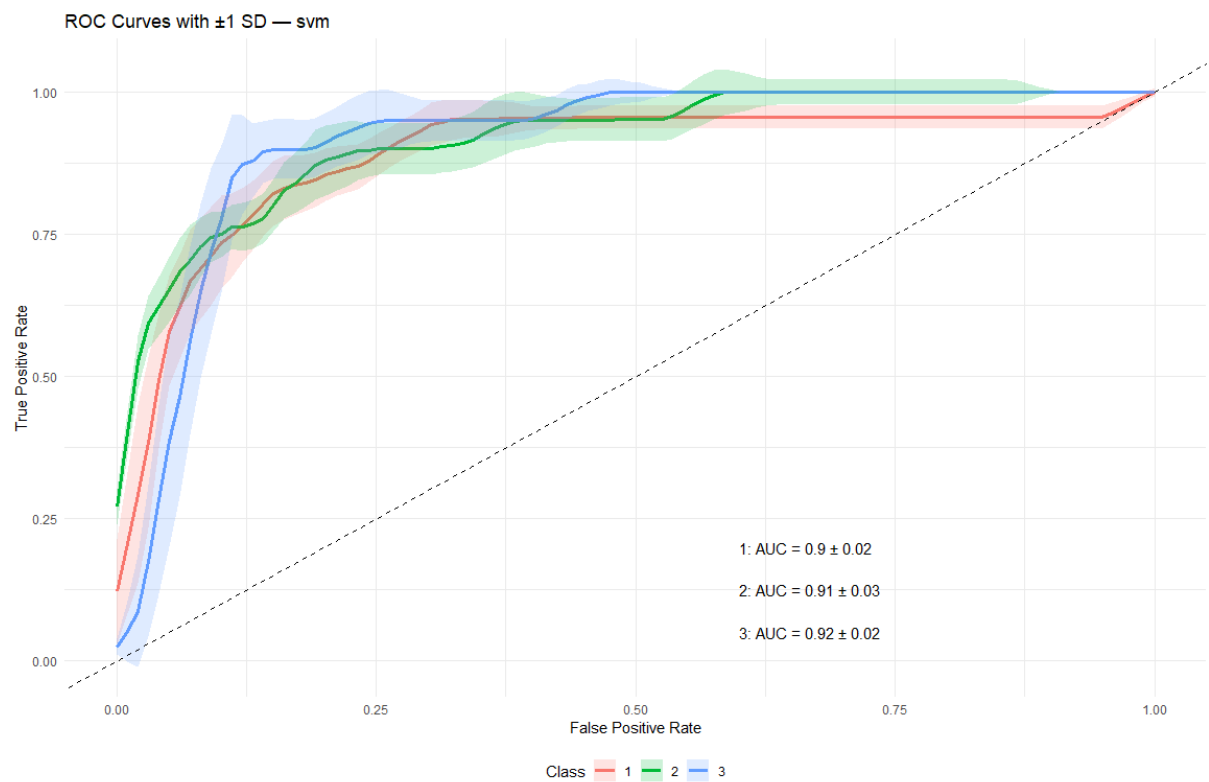
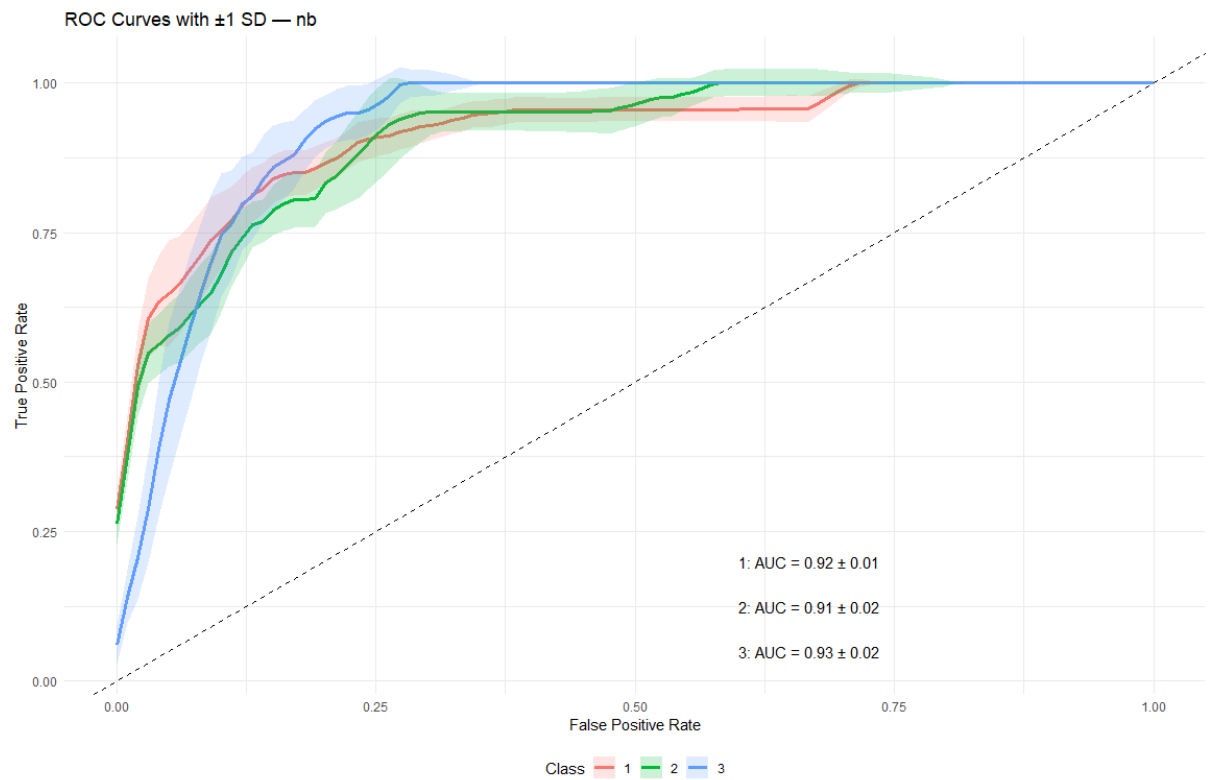
Wyniki

Wyniki dla zestawu 1:

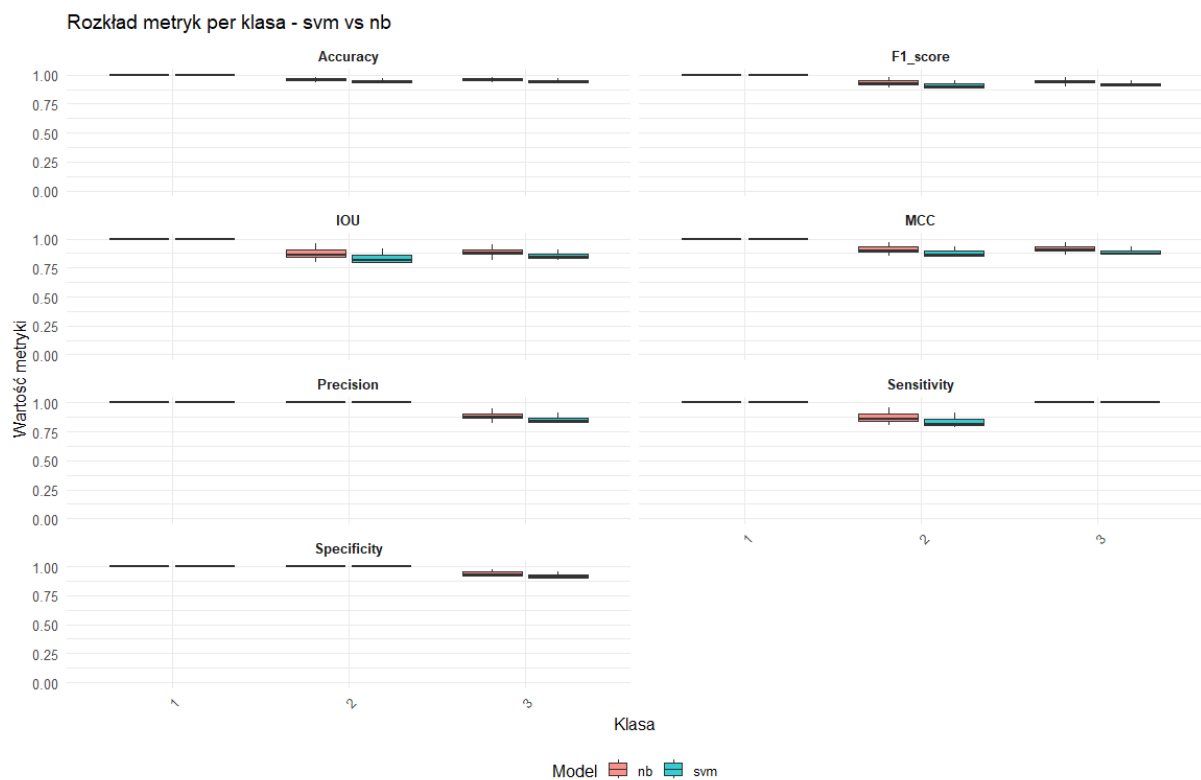
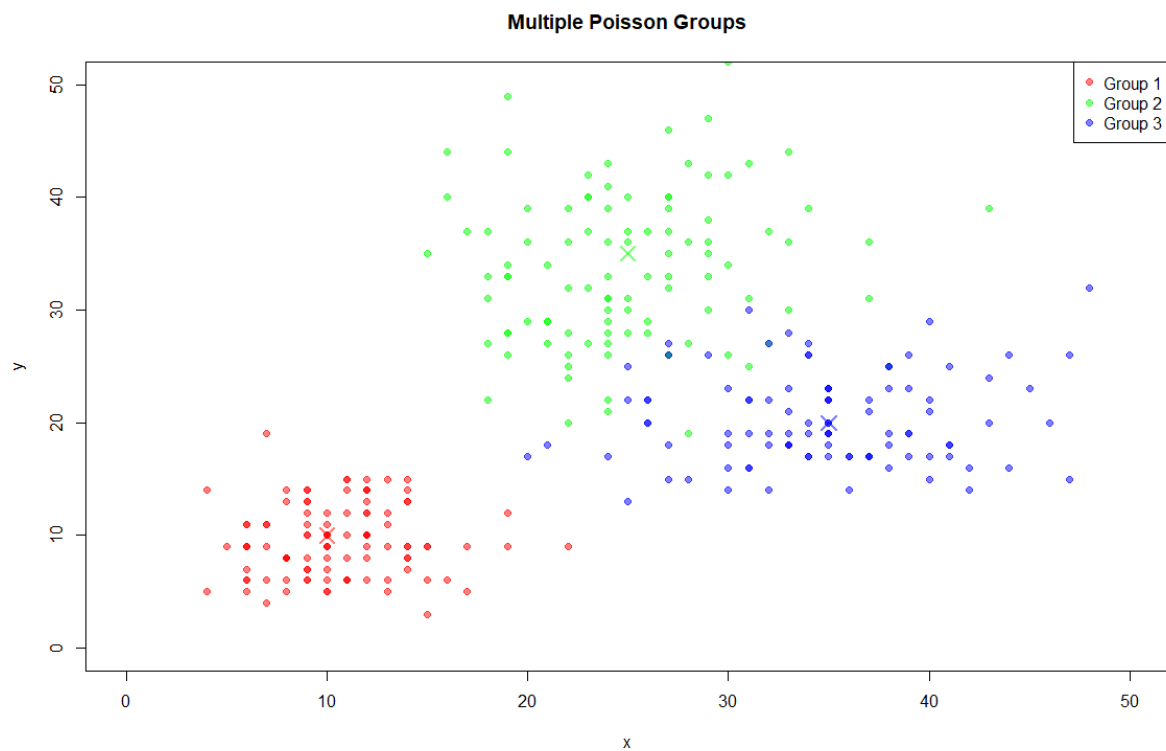


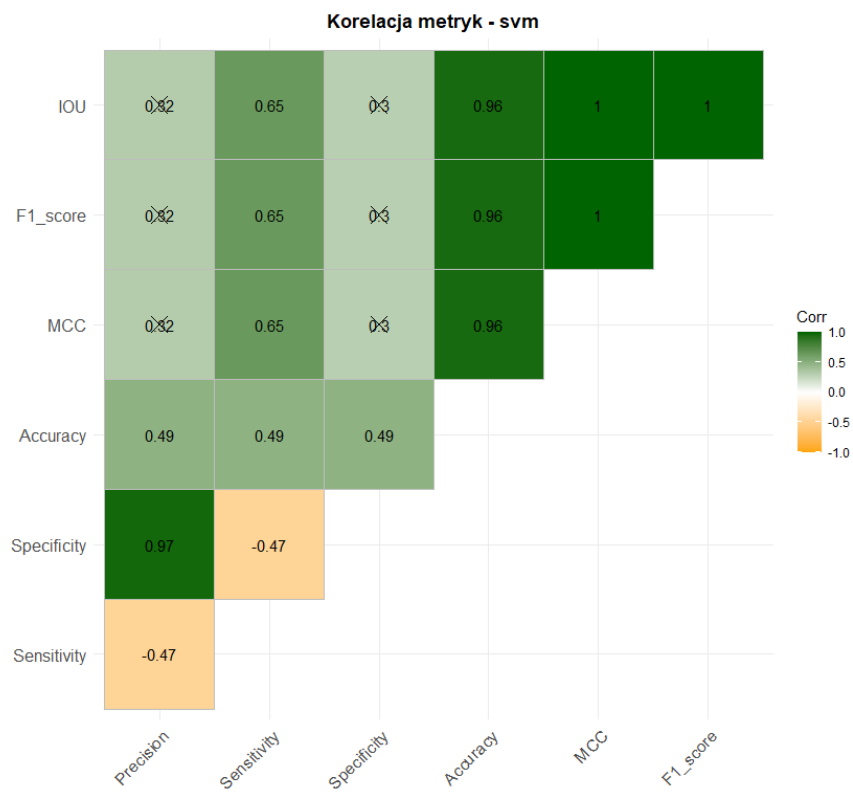
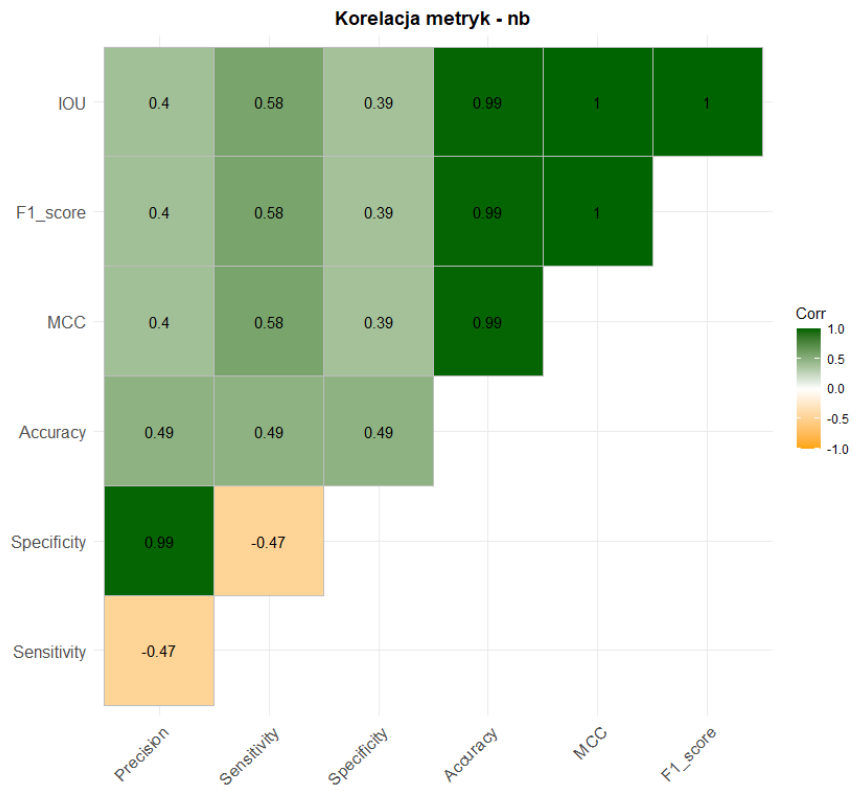


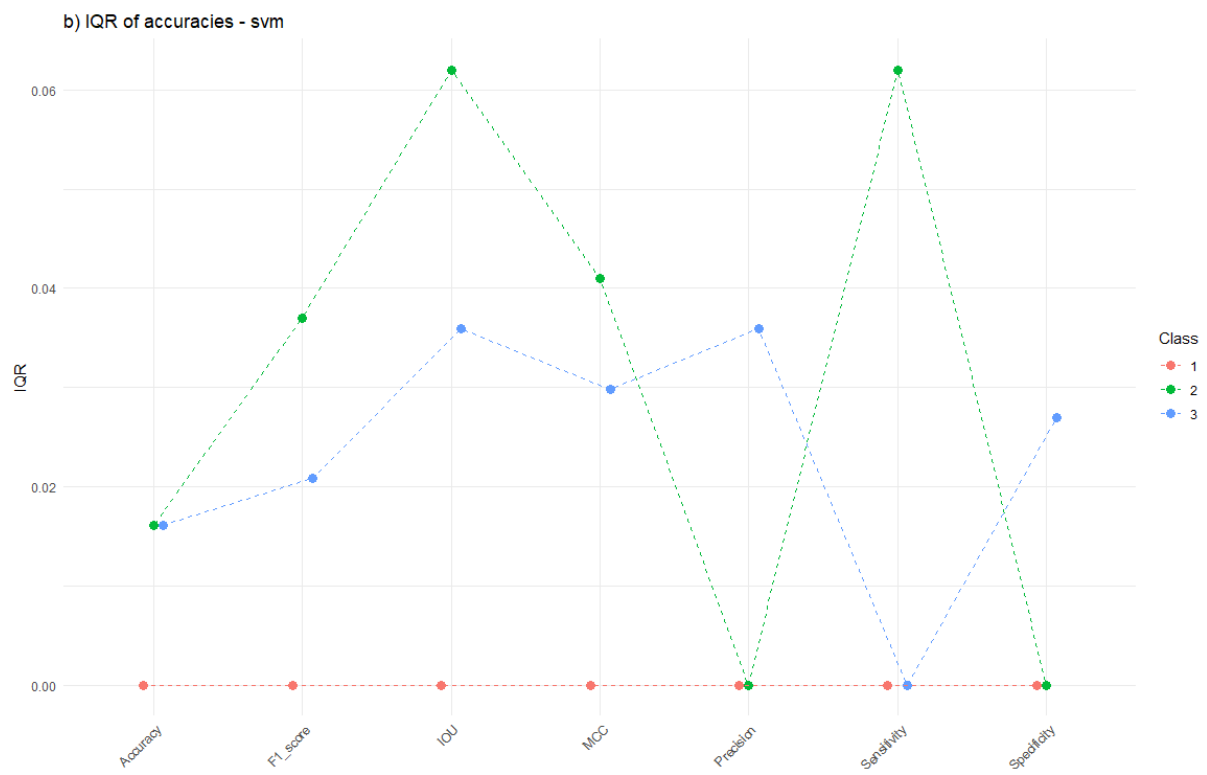
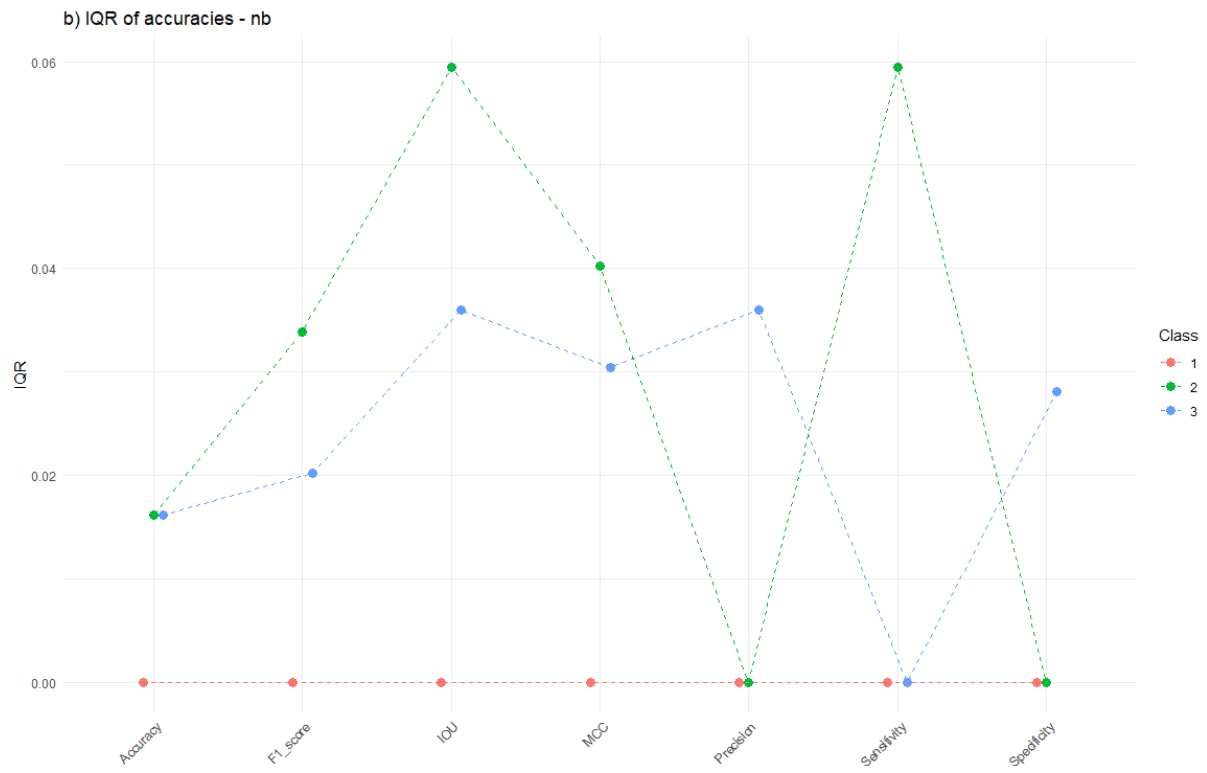


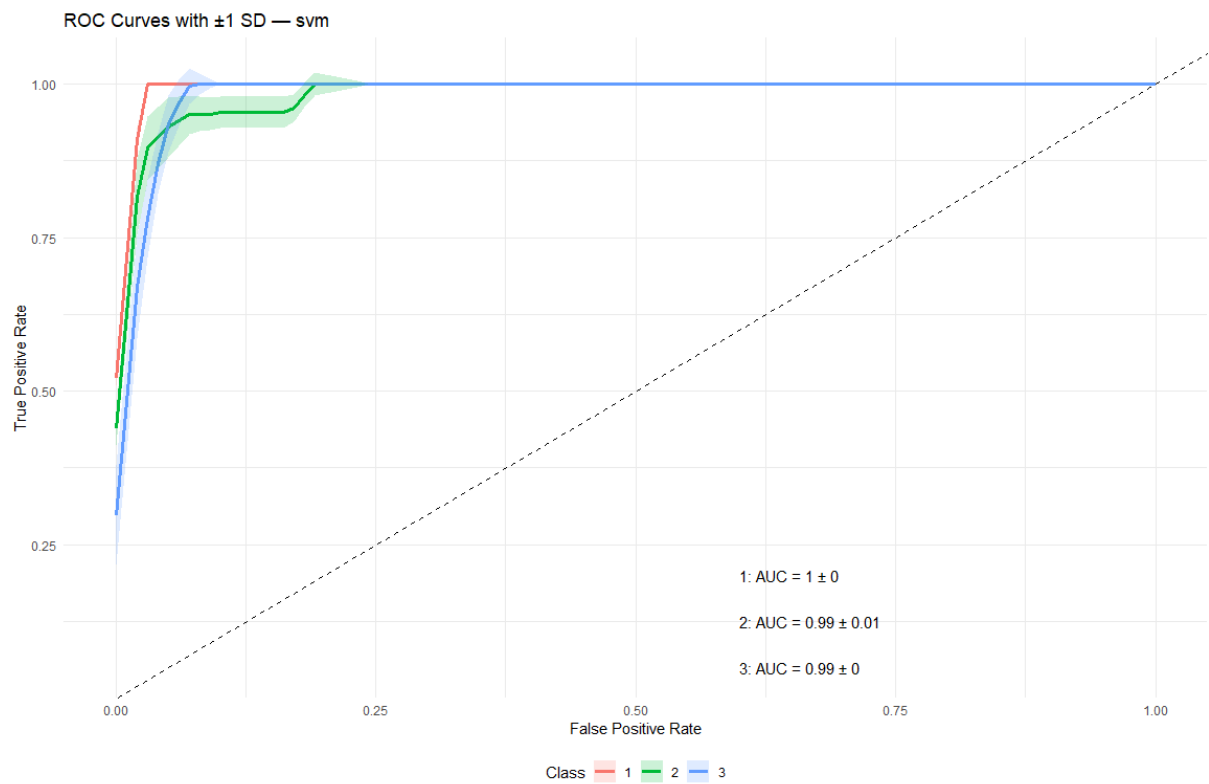
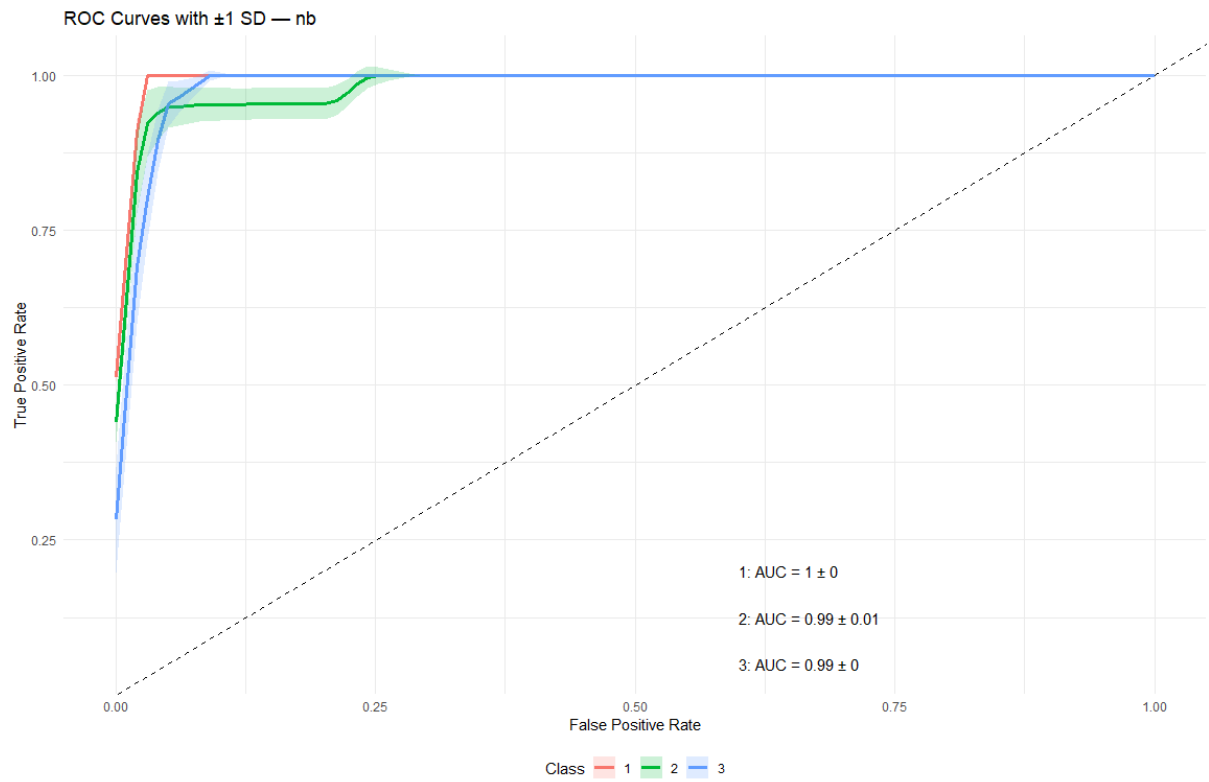


Wyniki dla zestawu 2:

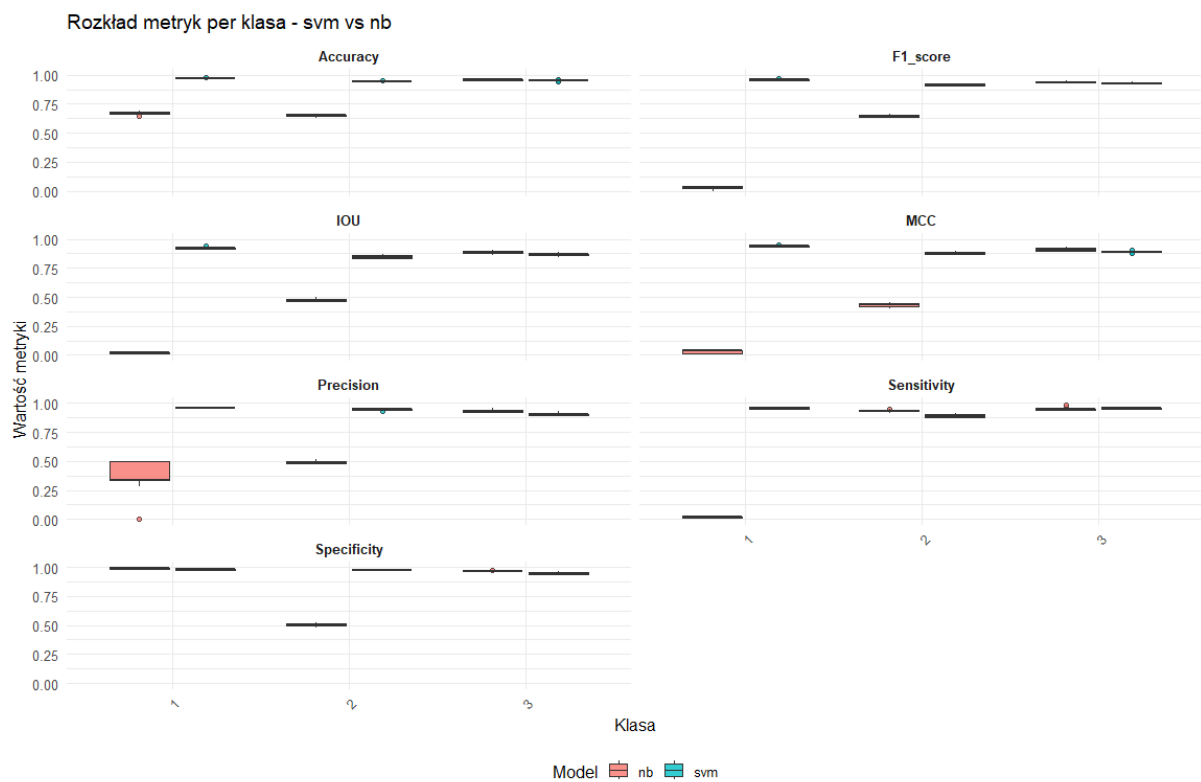
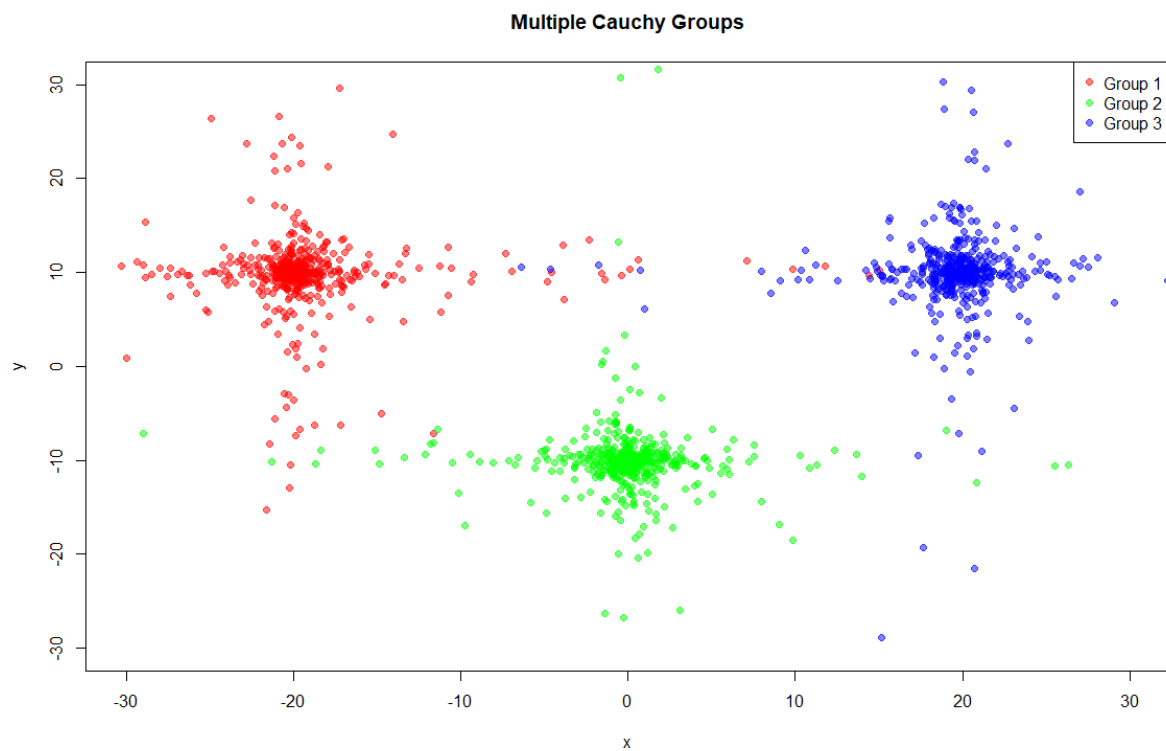


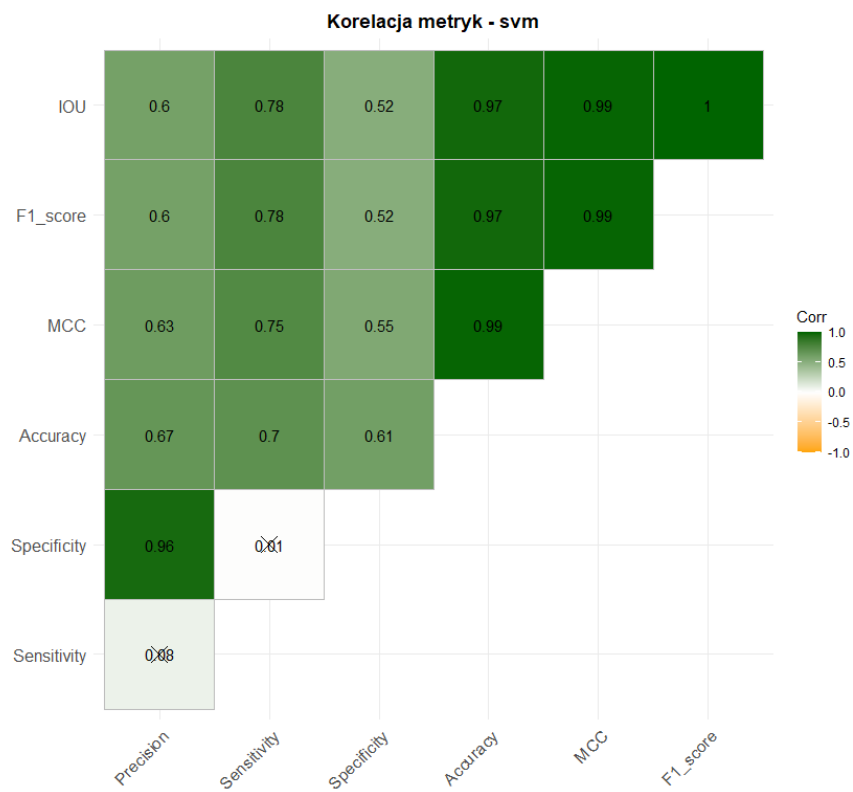
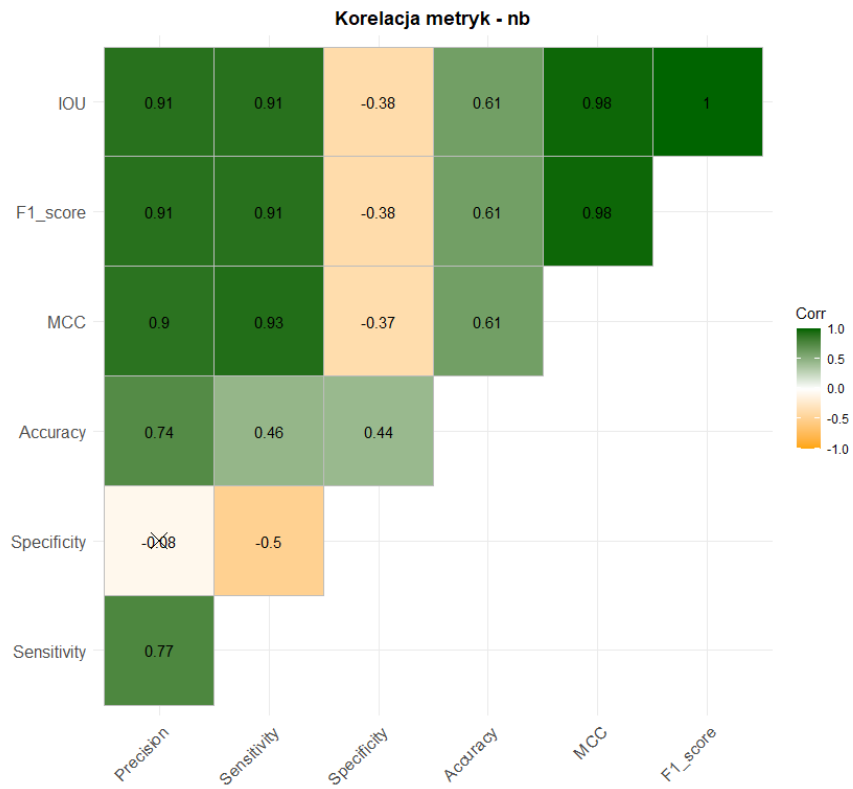


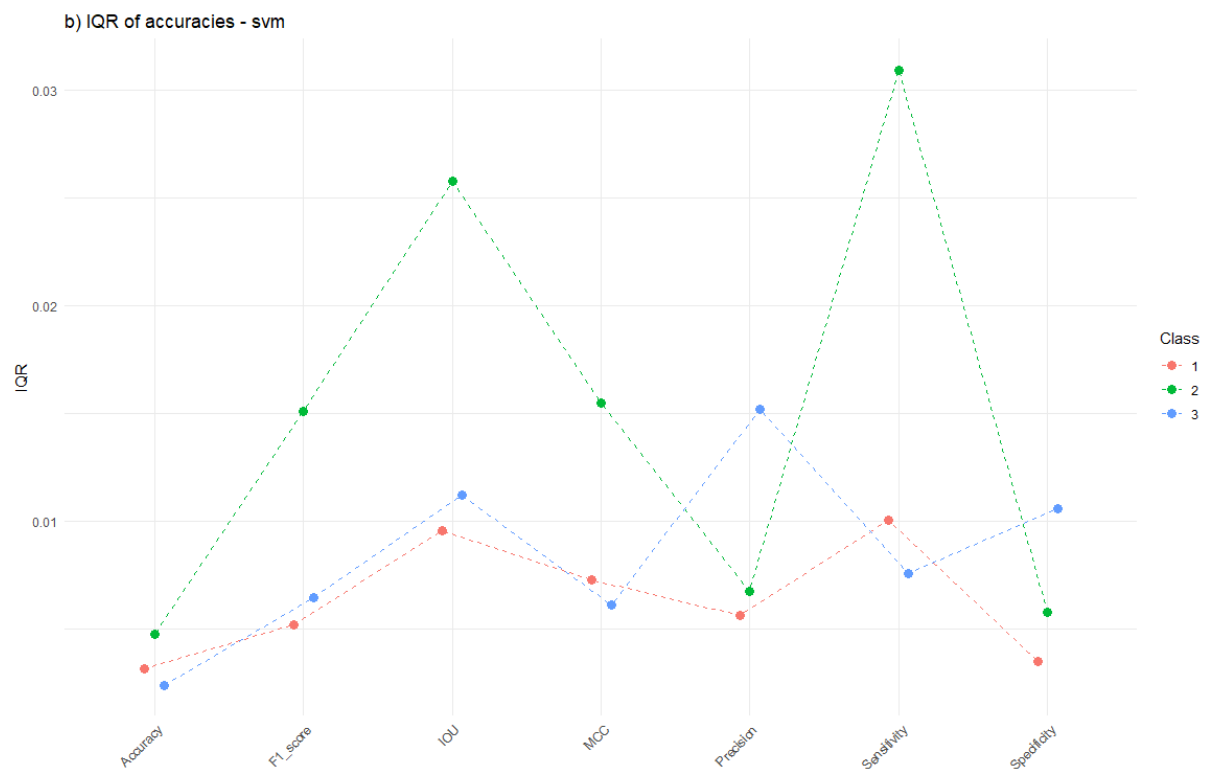
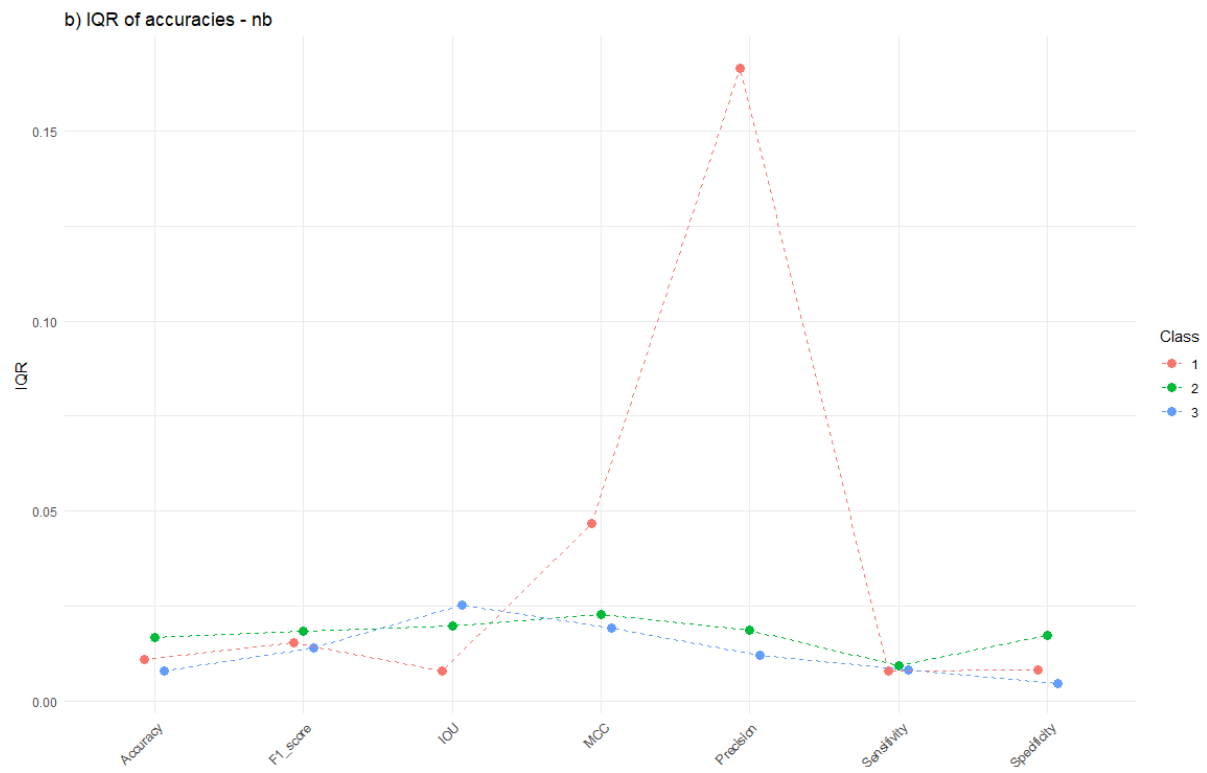


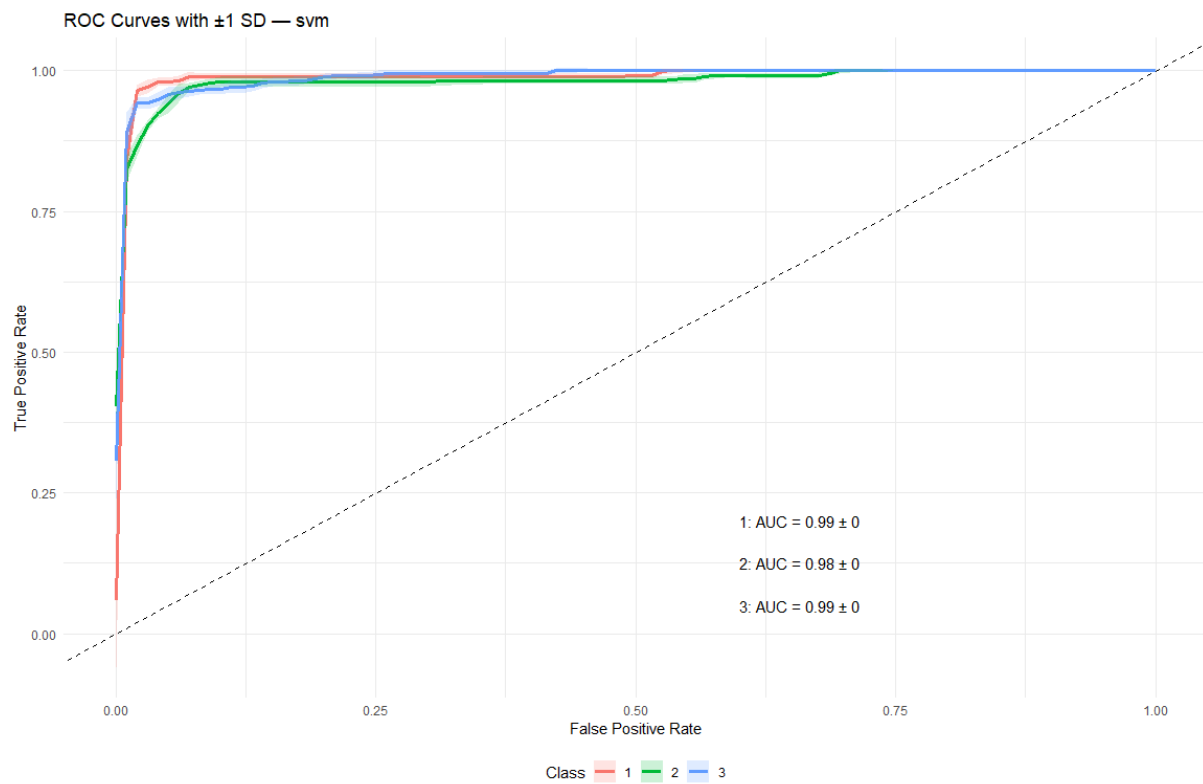
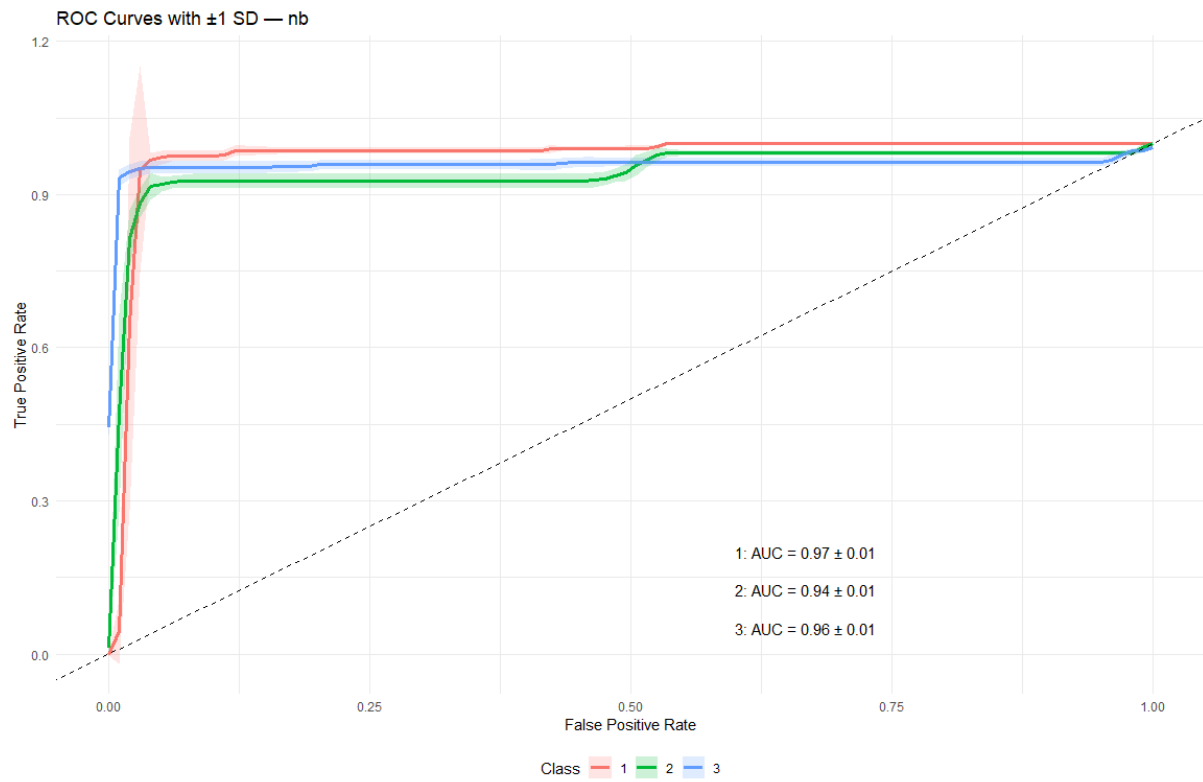


Wyniki dla zestawu 3:

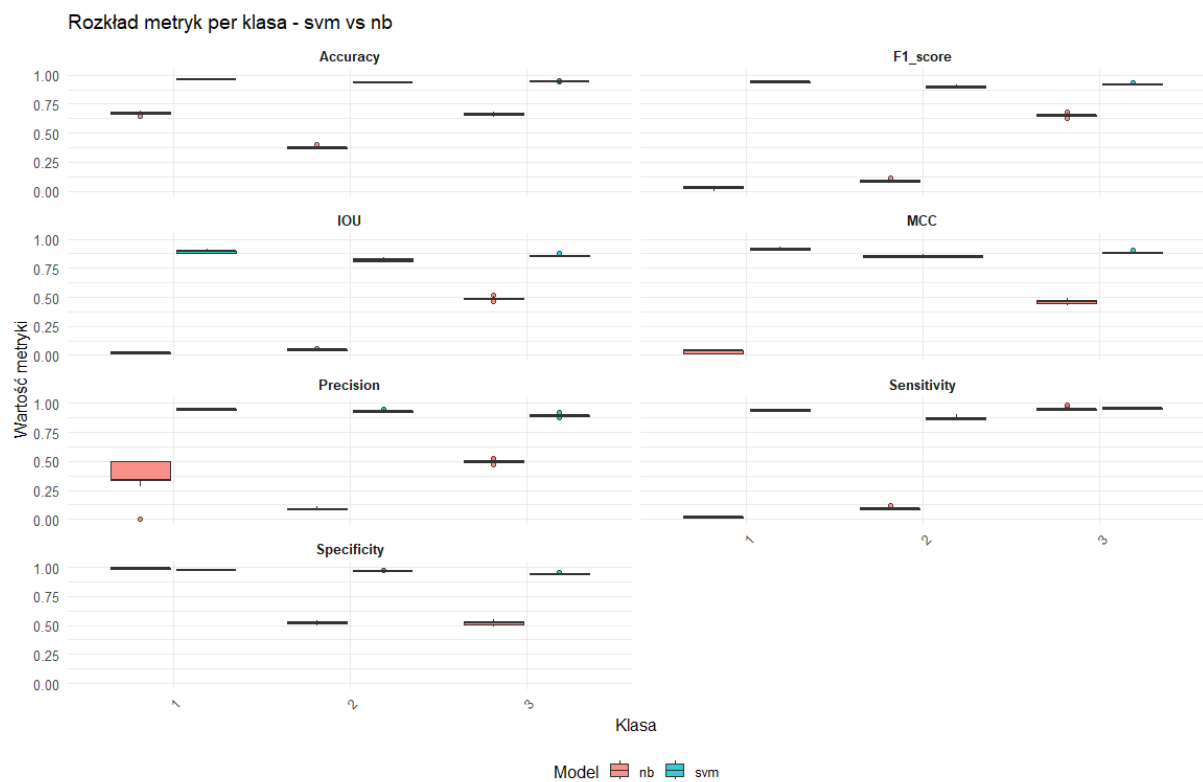
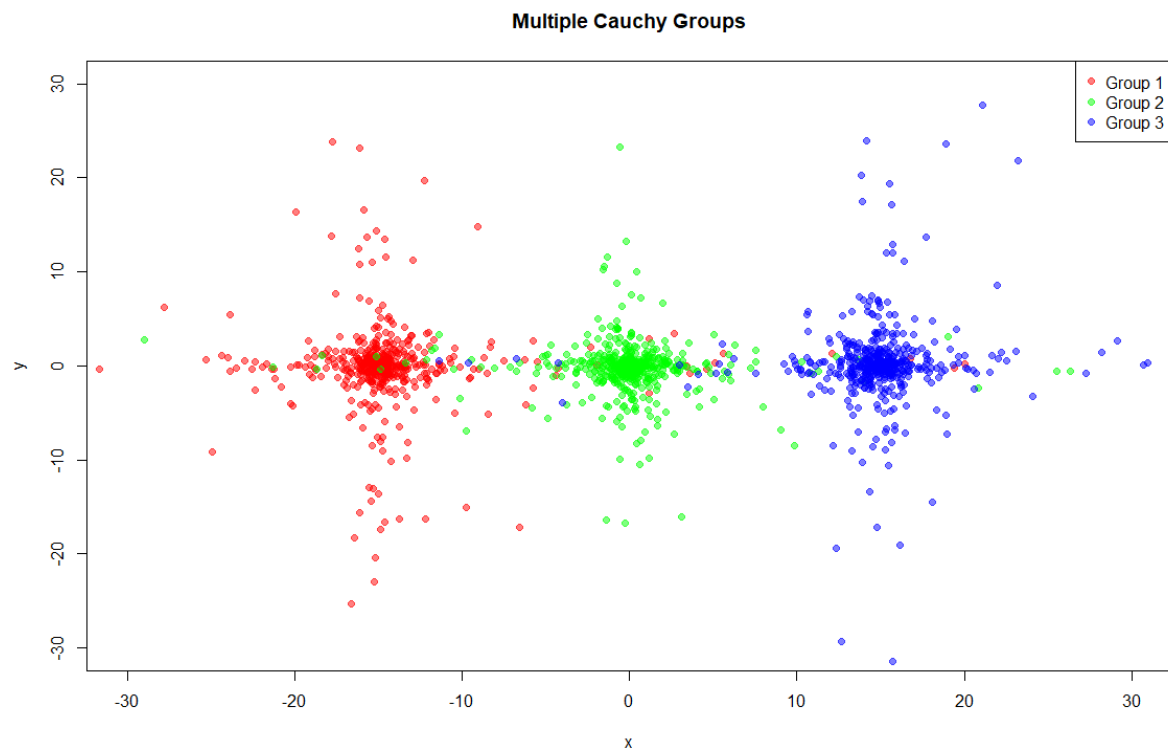


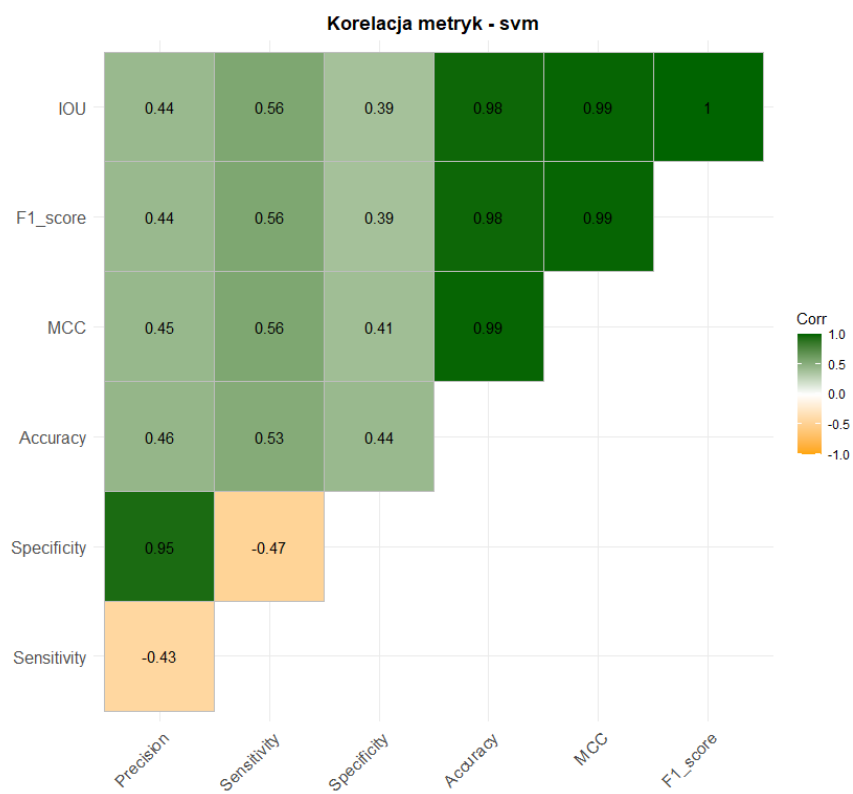
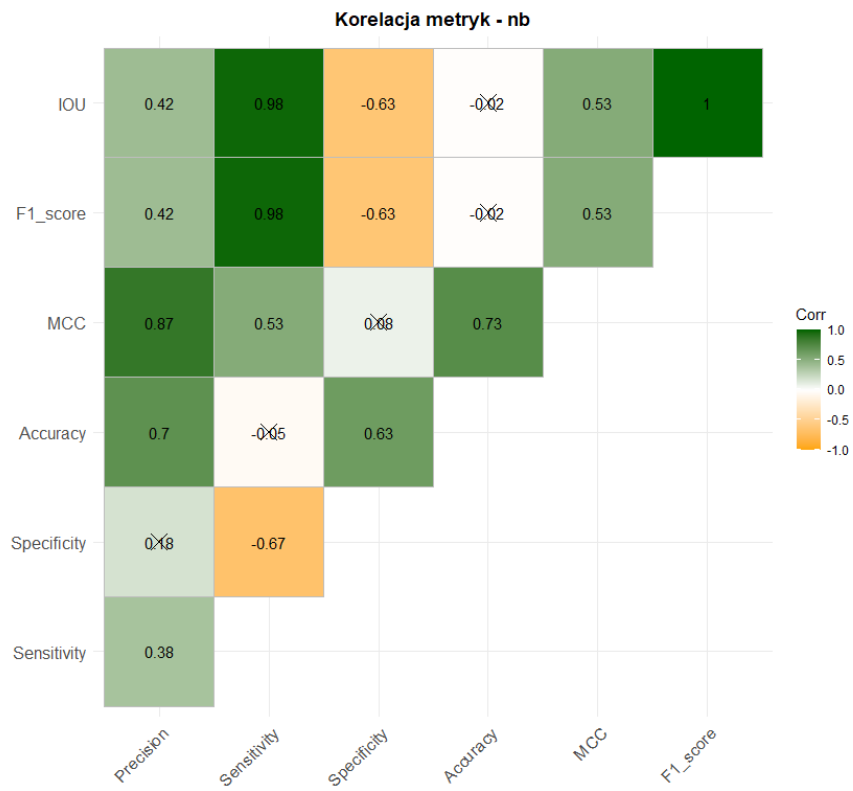




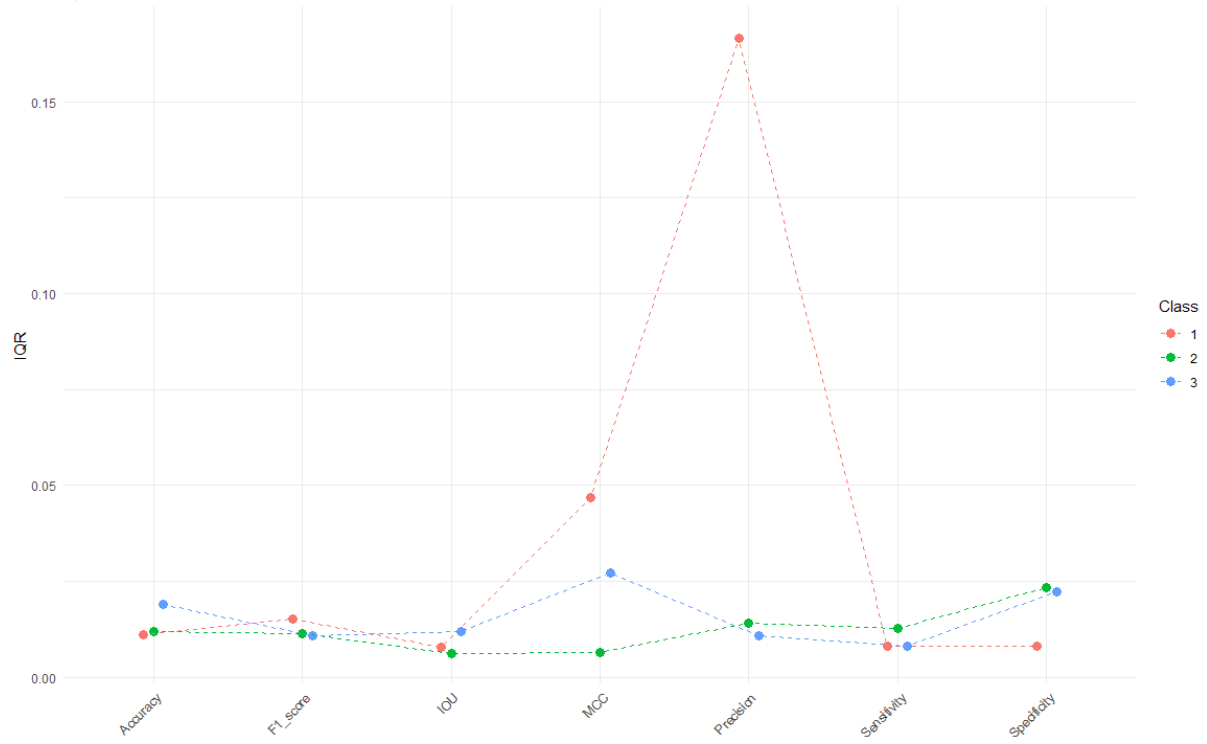


Wyniki dla zestawu 4:

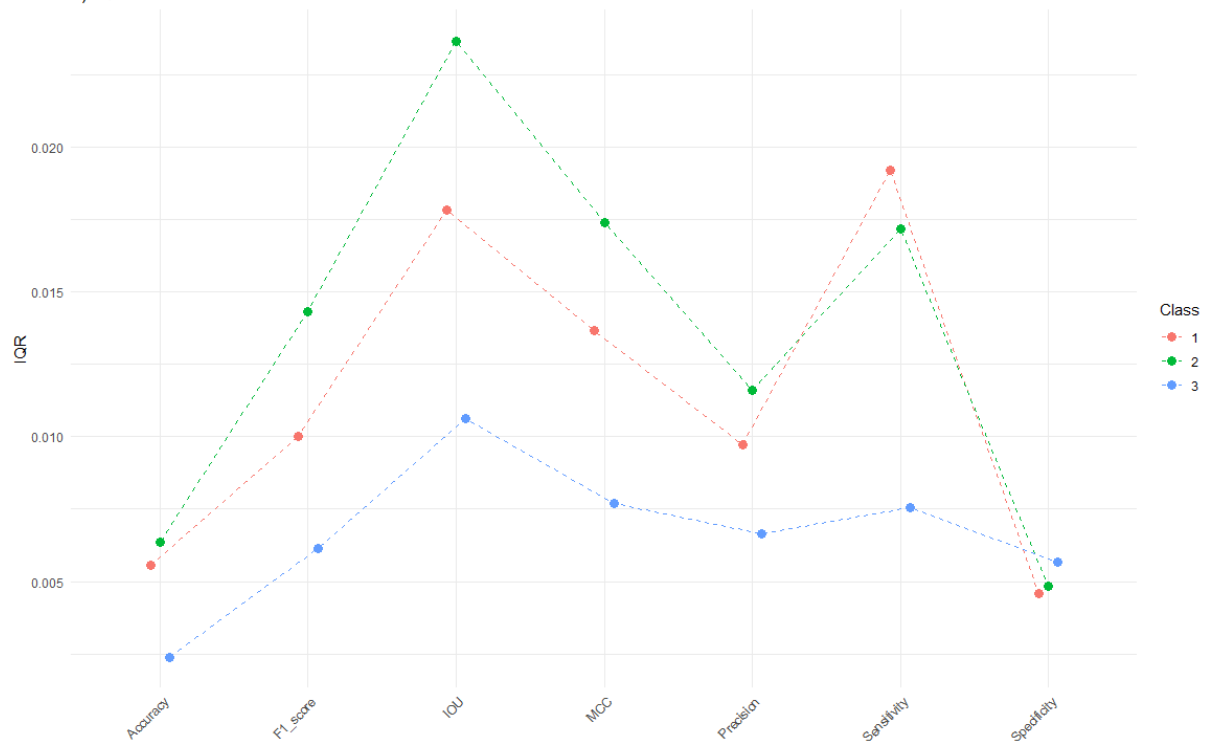


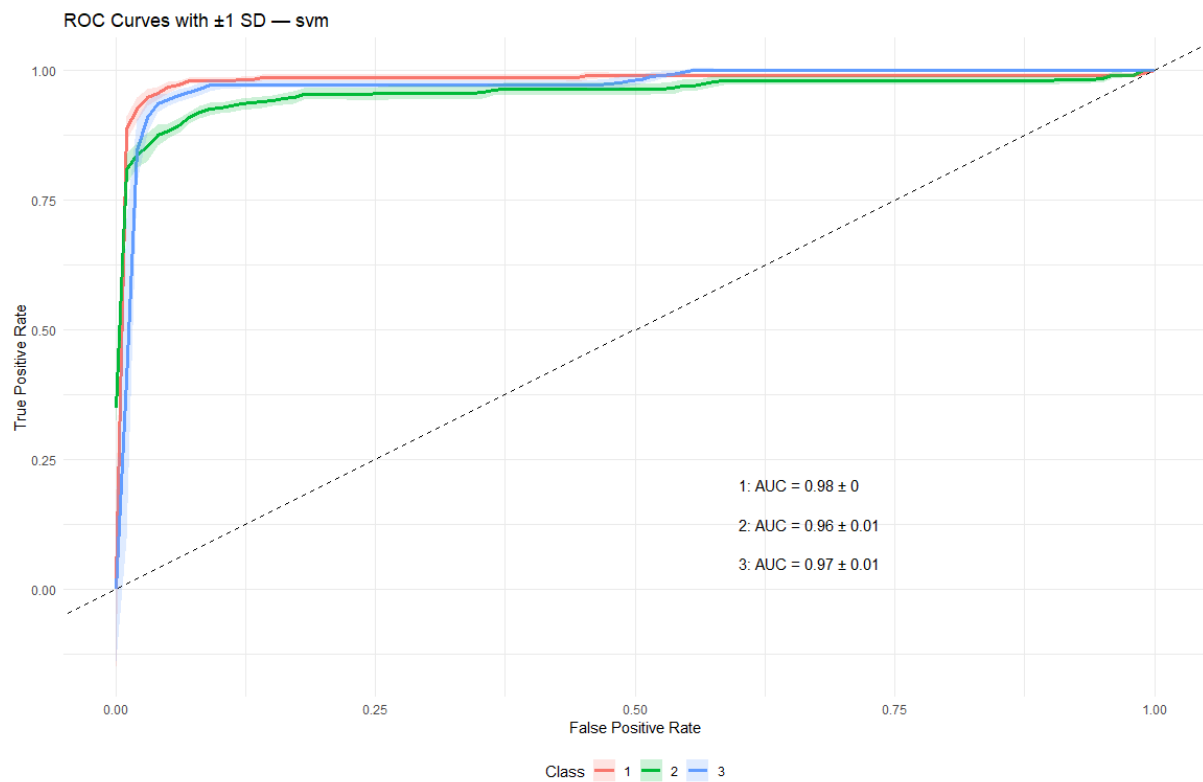
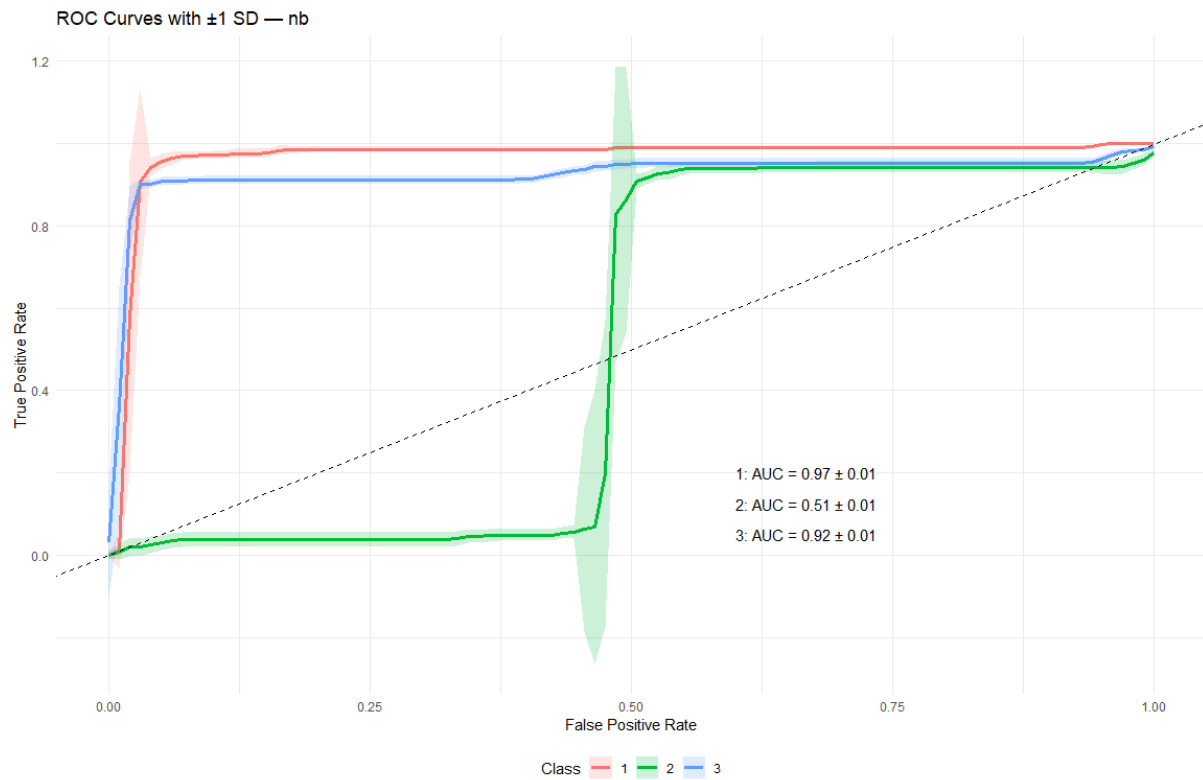


b) IQR of accuracies - nb

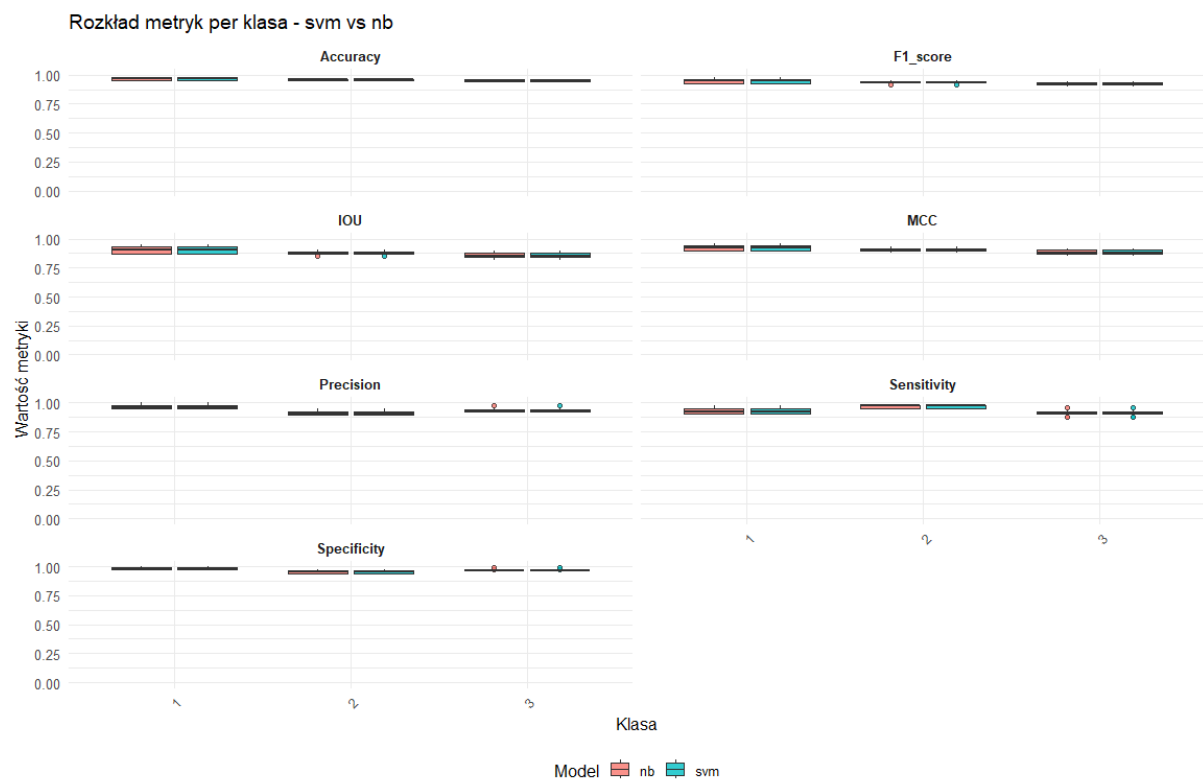
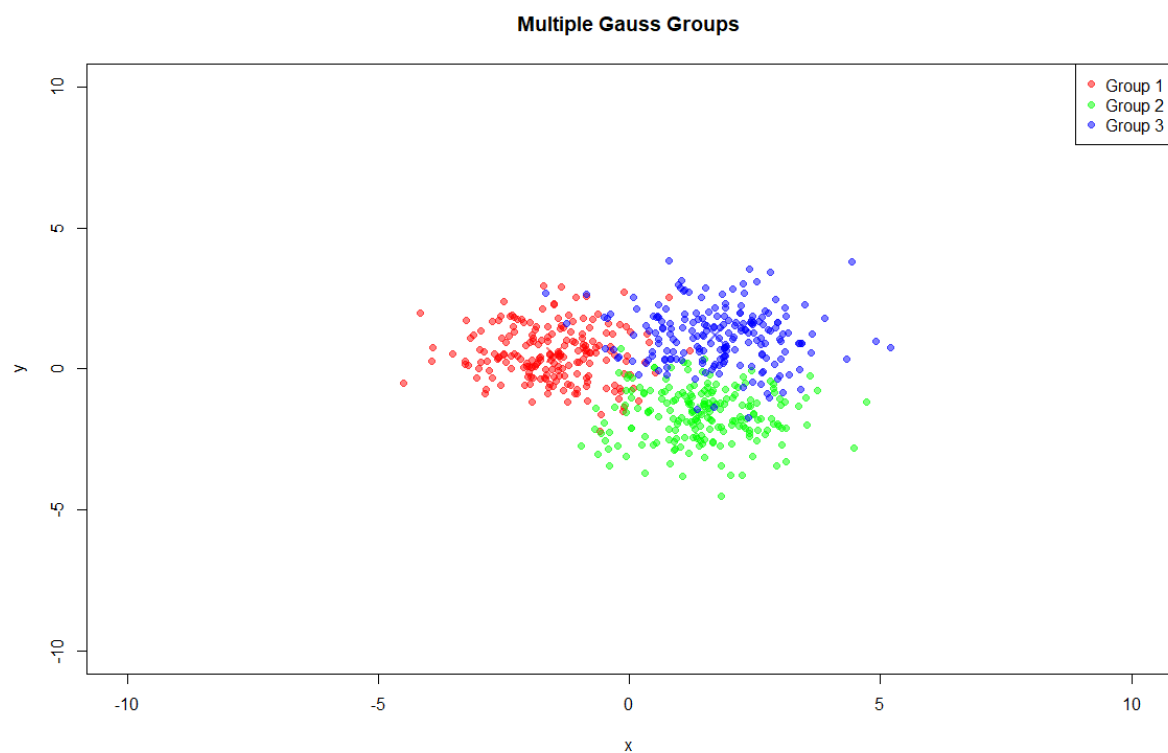


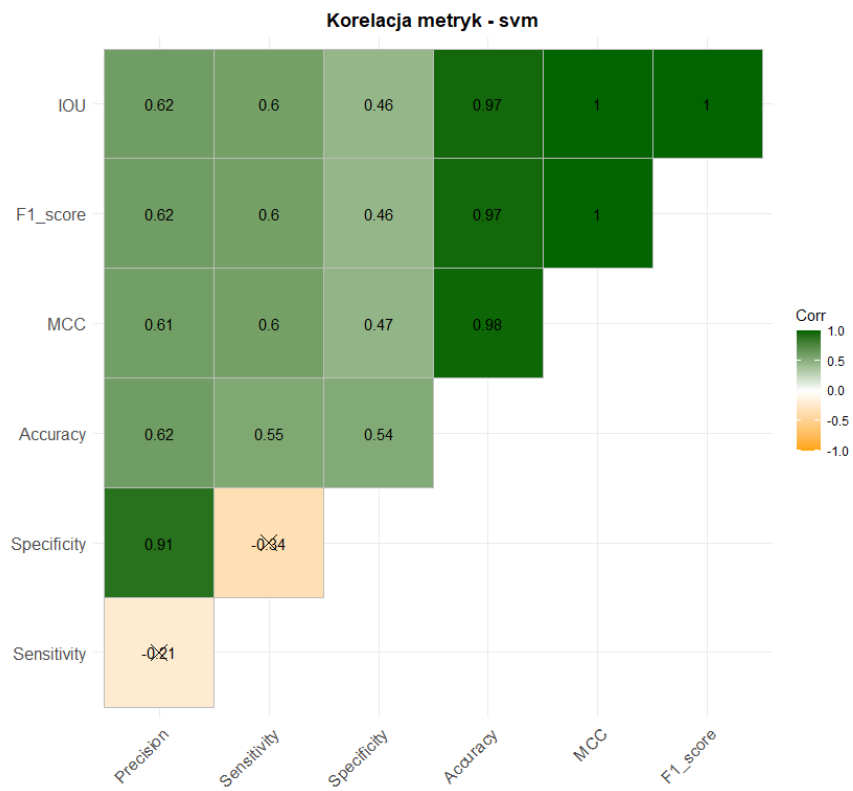
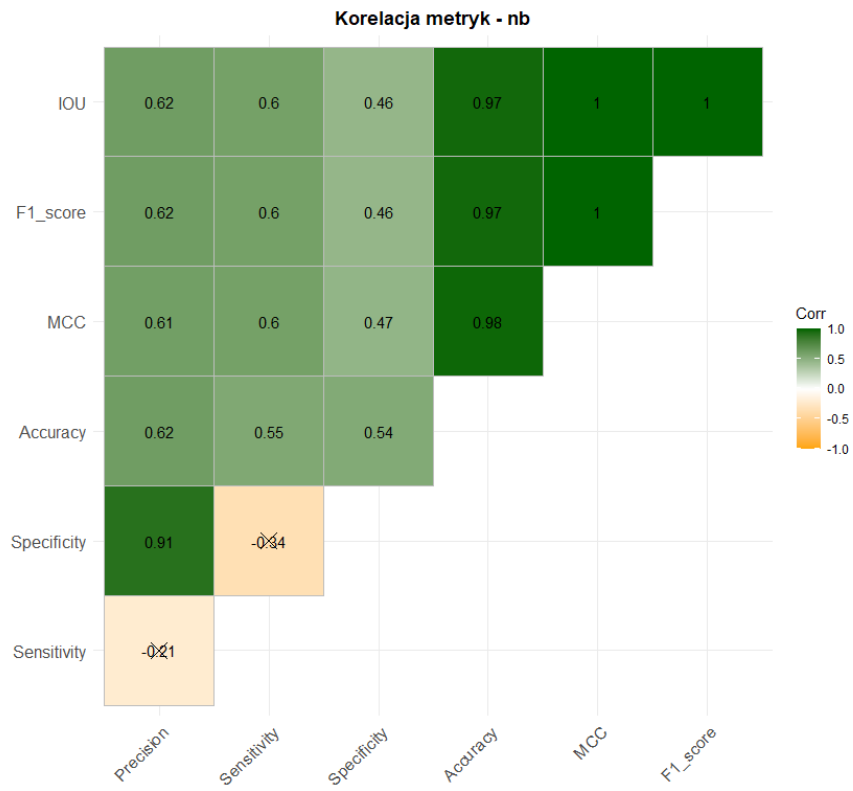
b) IQR of accuracies - svm



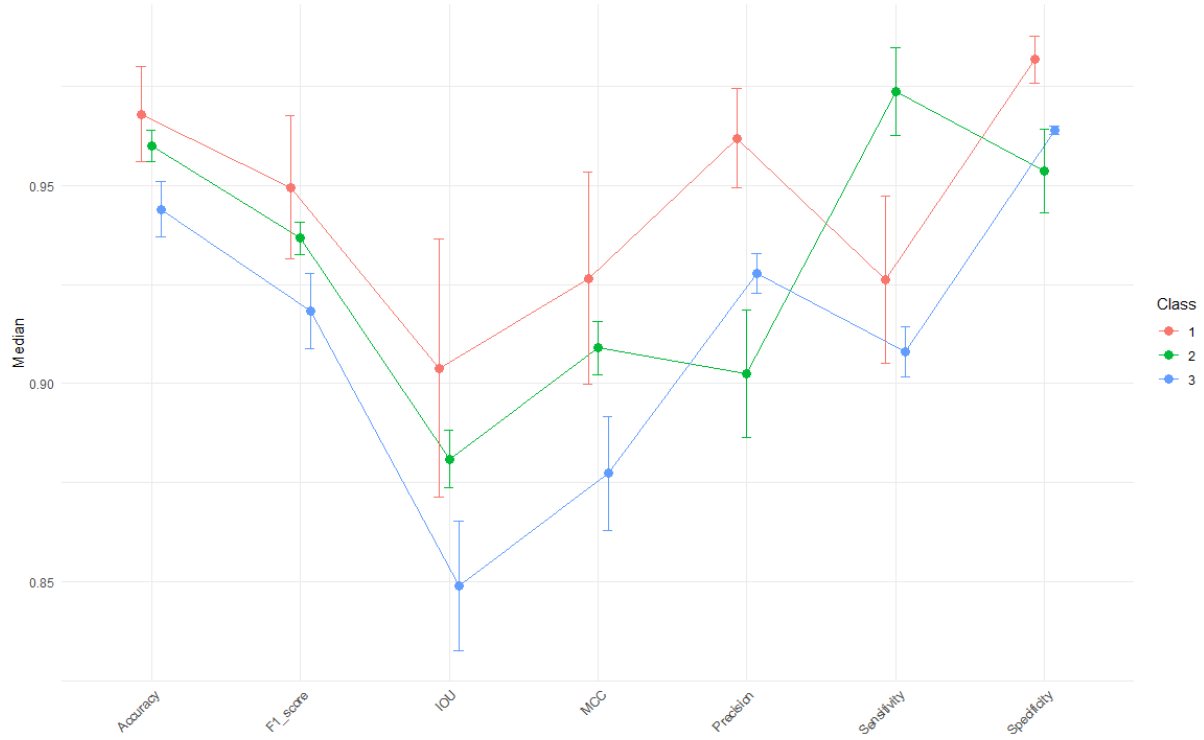


Wyniki dla zestawu 5:

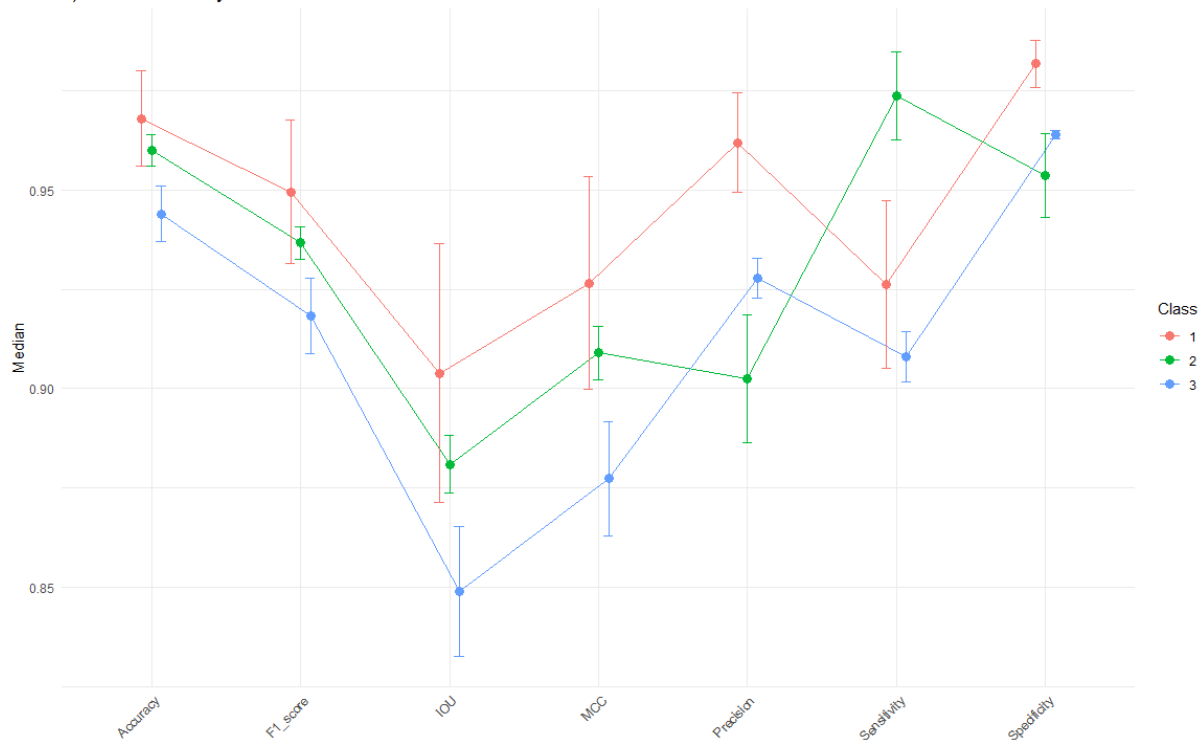


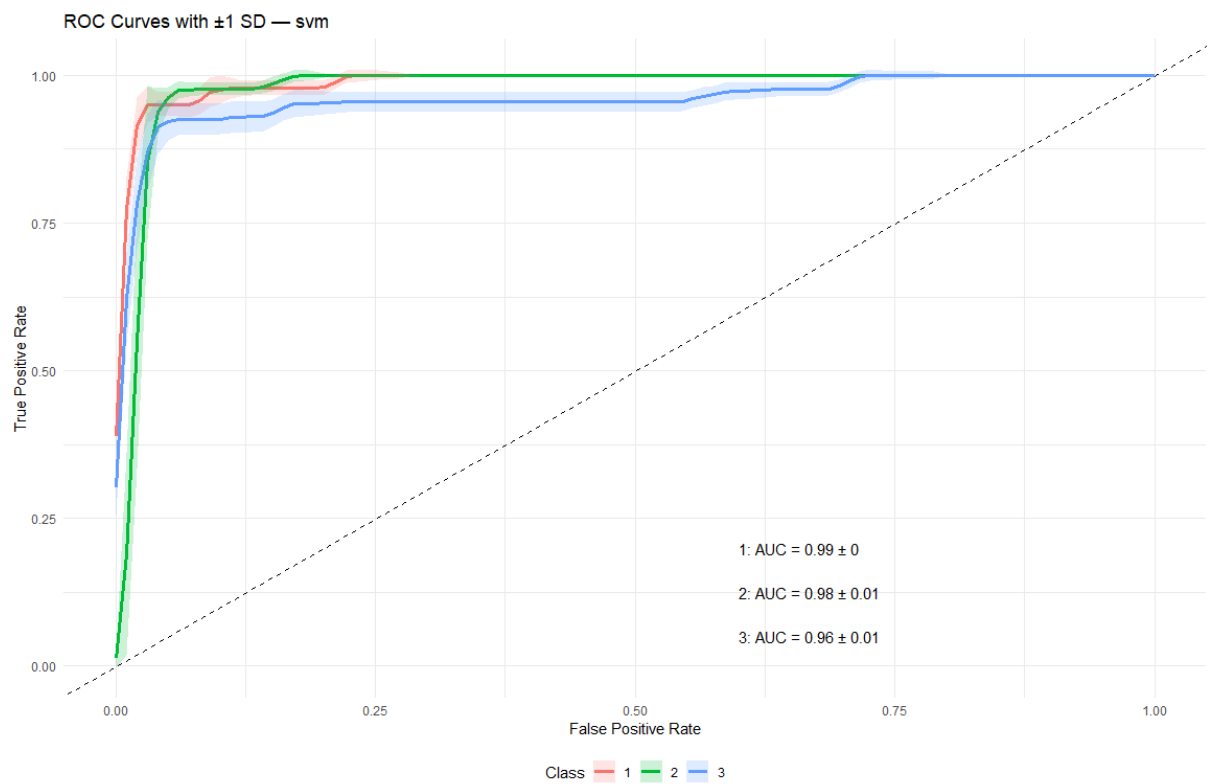
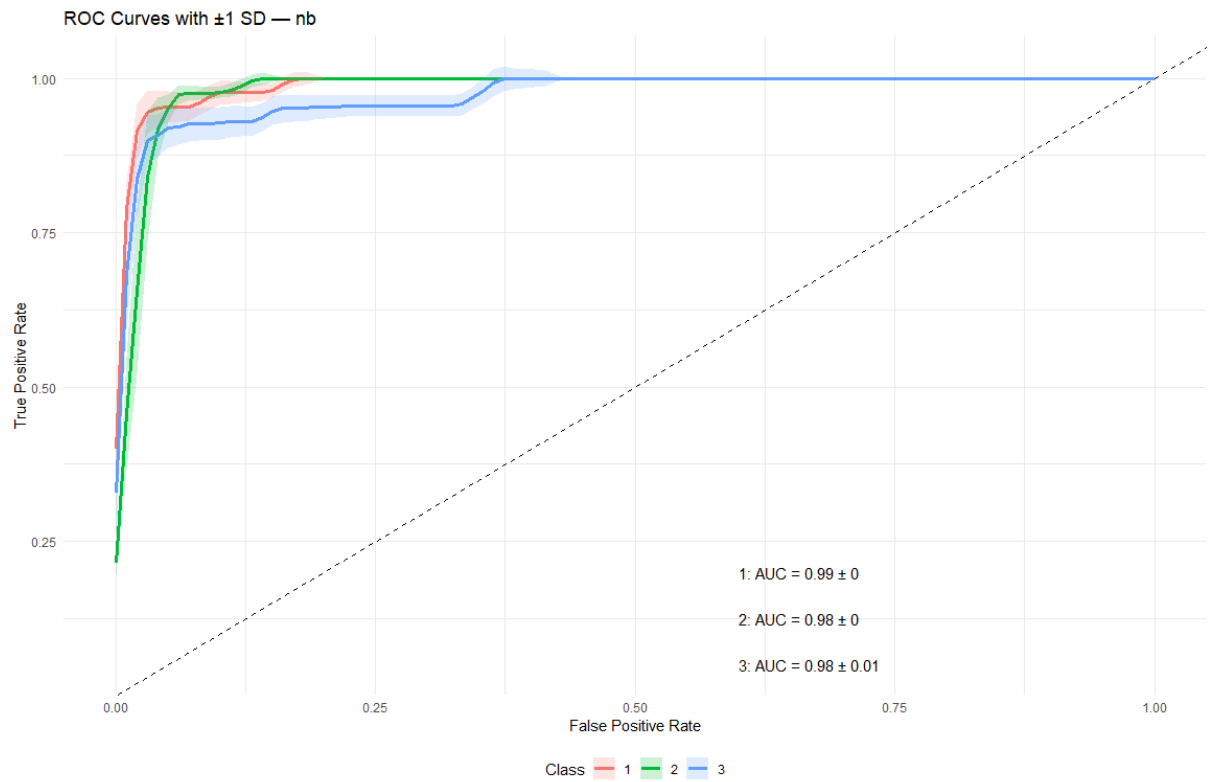


a) Median accuracy - nb

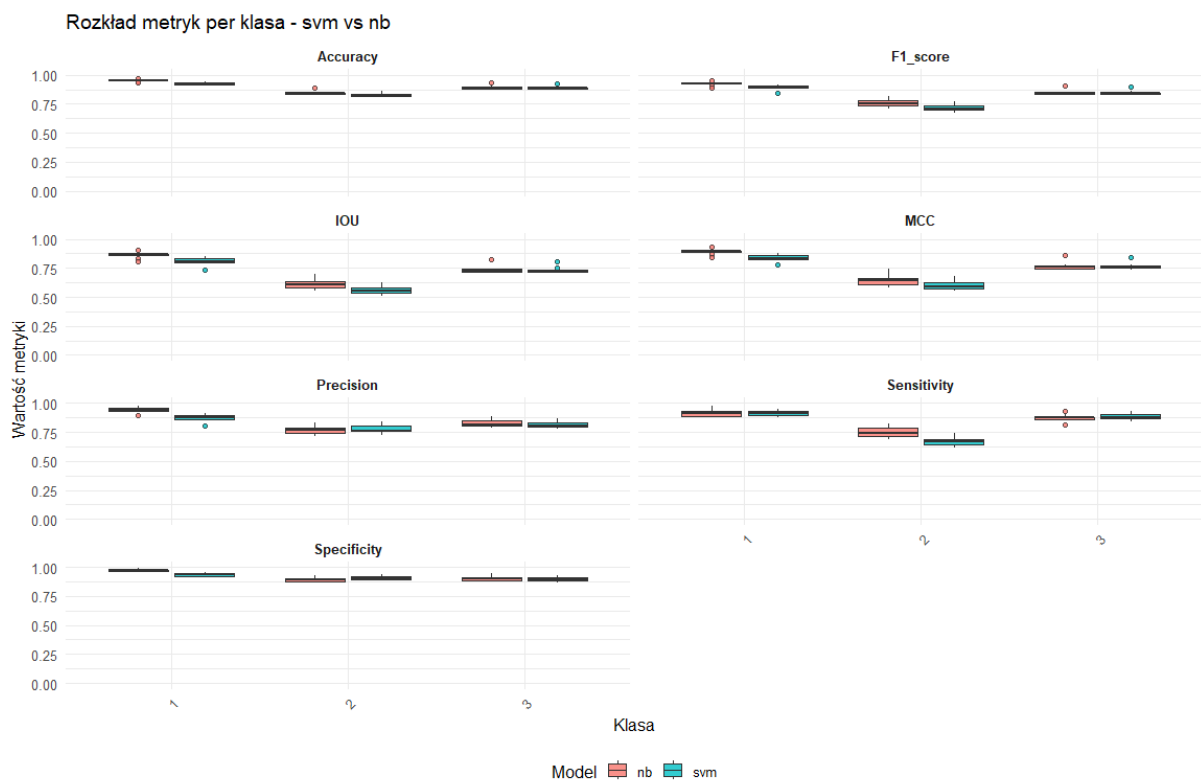
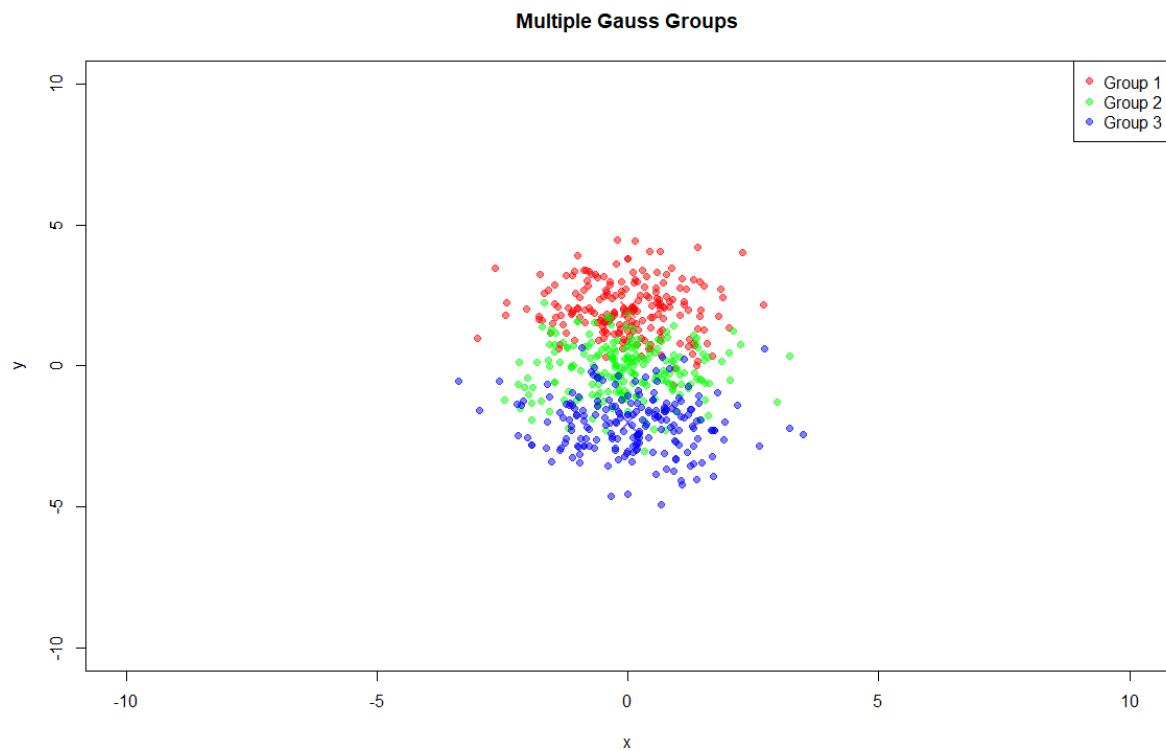


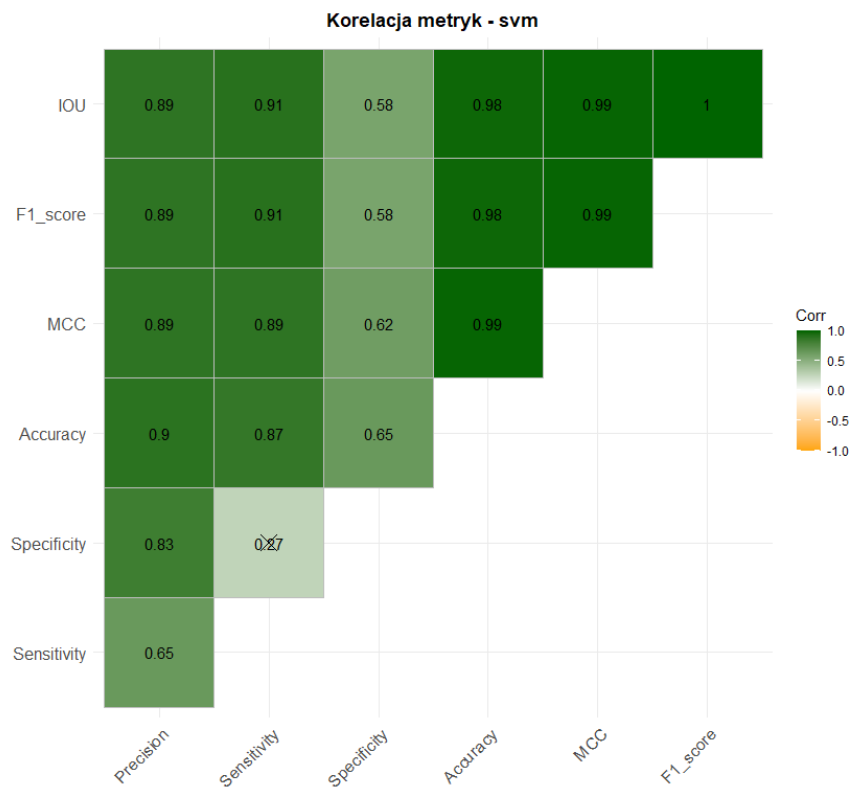
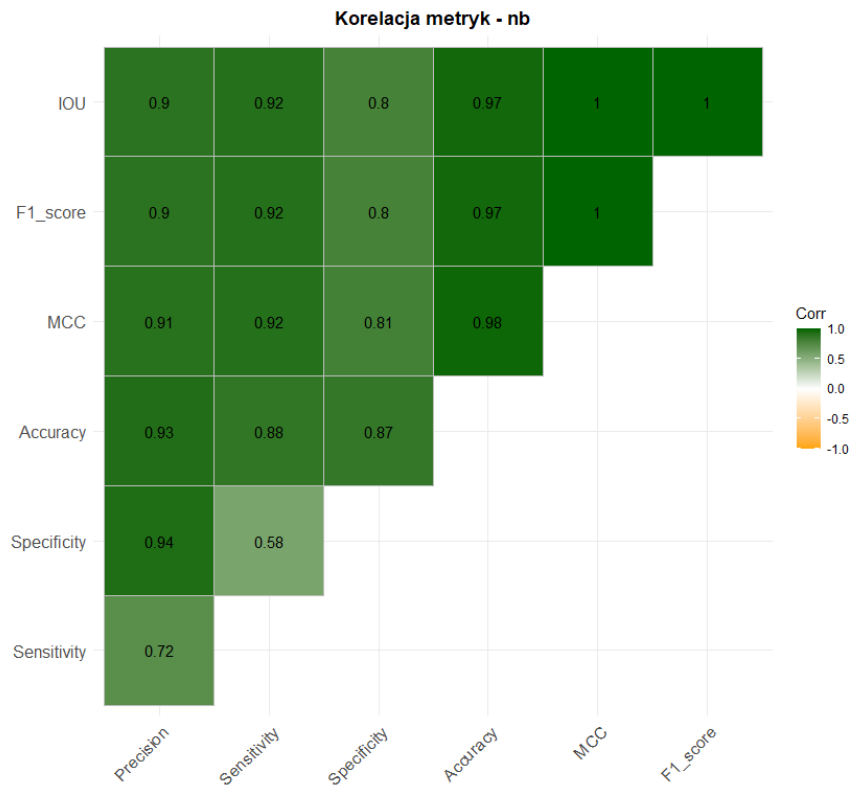
a) Median accuracy - svm

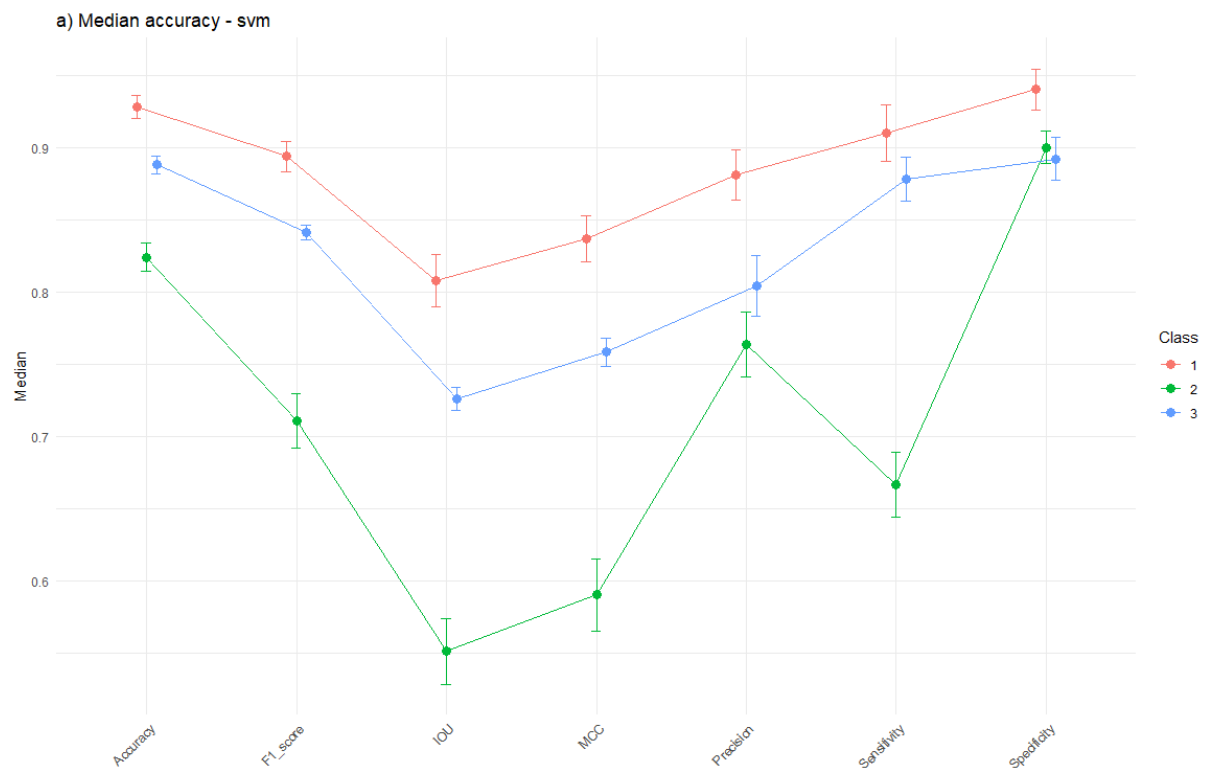
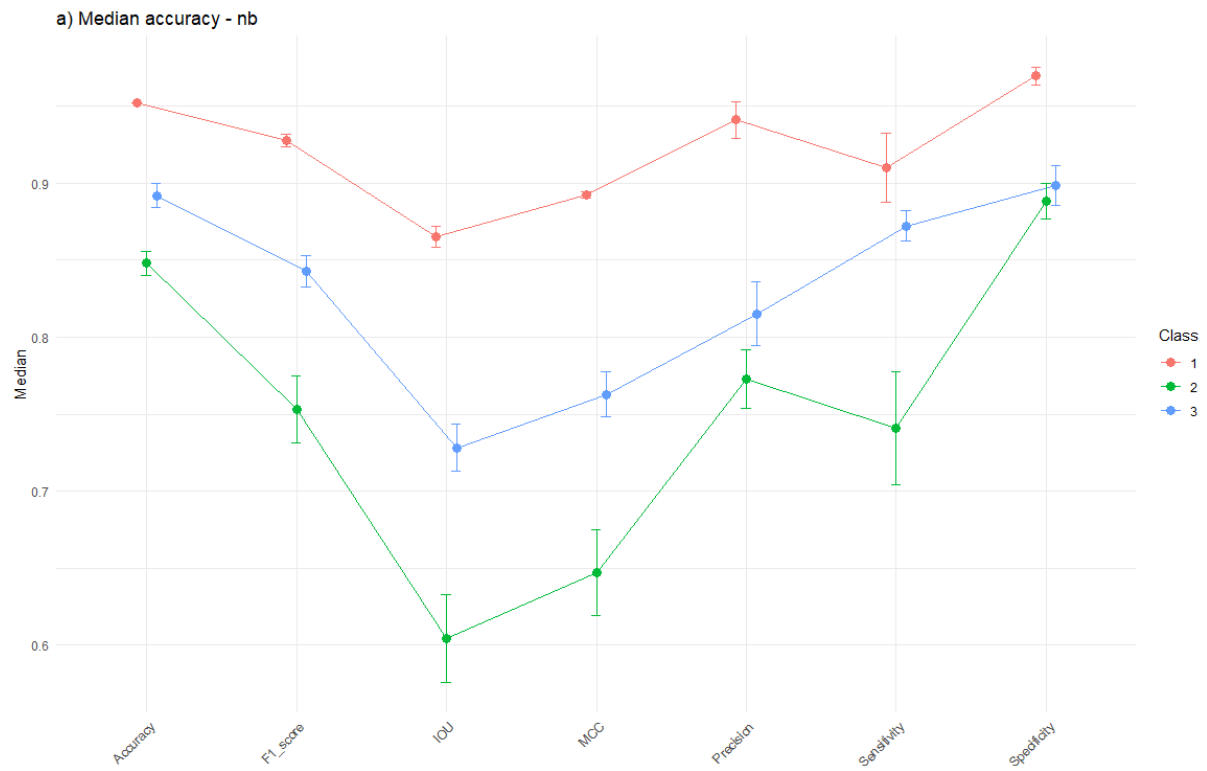


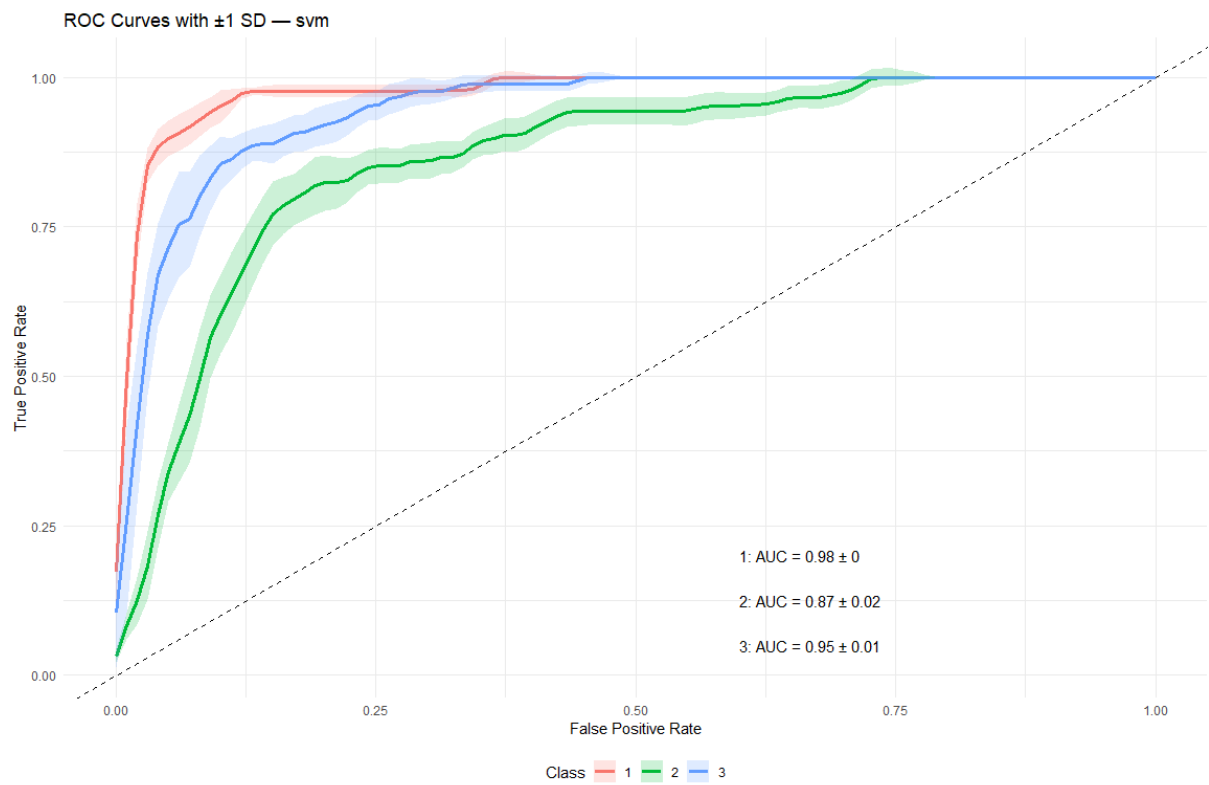
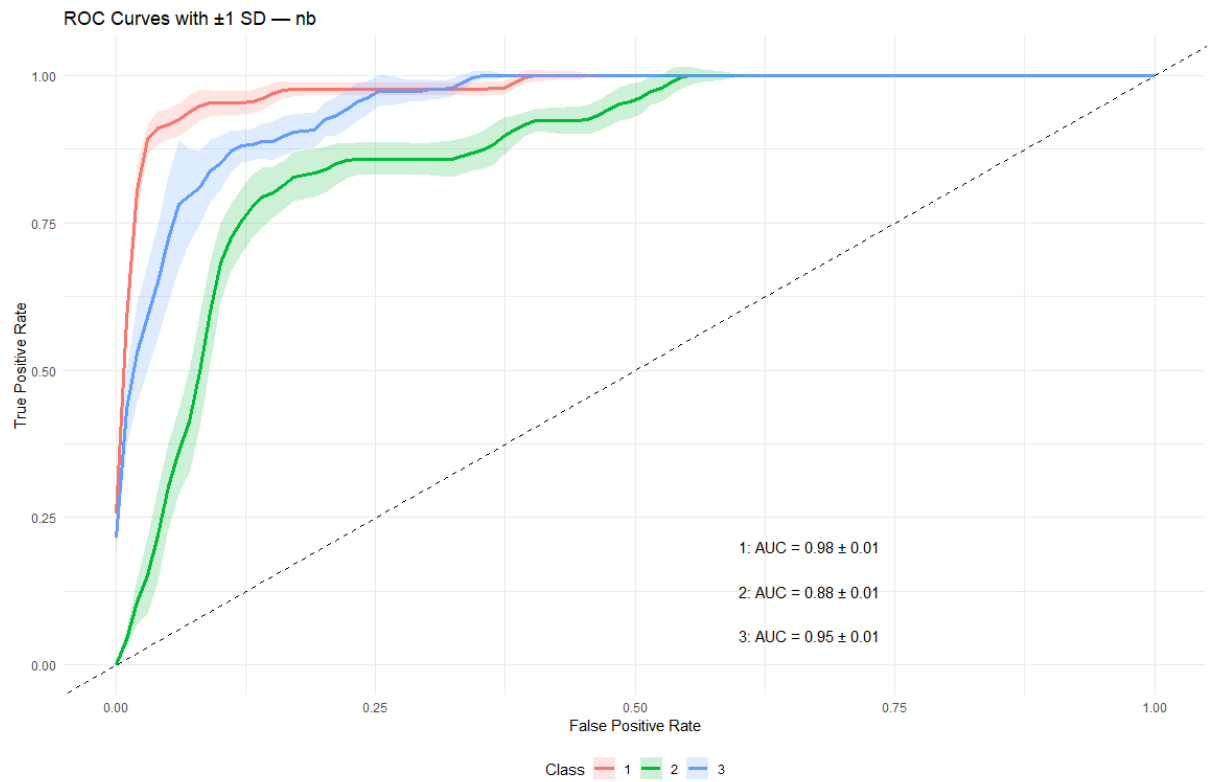


Wyniki dla zestawu 6:









Analiza wyników

W pierwszym zestawie danych, z mocno nachodzącymi na siebie grupami w dystrybucji Poisson'a, generalnie nieco lepiej sprawdził się klasyfikator SVM, szczególnie w przypadku grupy trzeciej. Klasyfikacja takich przypadków jak pierwszy zestaw danych stanowi duże wyzwanie ze względu na wysoki stopień pokrycia pomiędzy grupami.

W drugim zestawie – dystrybucja Poisson'a z częściowym pokryciem grupy 2 i 3 oraz mocno oddaloną grupą 1 – możemy na pierwszy rzut oka zaobserwować nieznacznie większą skuteczność naiwnego klasyfikatora Bayes'a. Oba klasyfikatory uzyskały celność bliską 100% dla grupy 1 oraz bardzo dobre wyniki dla pozostałych dwóch grup. Wynika to z niskiego pokrycia pomiędzy grupami w zestawie testowym.

W trzecim zestawie, zawierającym mocno oddzielone od siebie grupy w dystrybucji Cauchy'ego, zdecydowanie lepszy wynik osiągnął klasyfikator SVM, przede wszystkim dla grup 1 oraz 2. W przypadku grupy 3 wynik jest mocno zbliżony. Na wykresach IQR możemy zauważyć wyjątkowo wysoki Interquartile Range w metryce Precision dla naiwnego klasyfikatora Bayes'a.

W czwartym zestawie, zawierającym horyzontalnie wyrównane zbiory o dystrybucji Cauchy'ego, ponownie znacznie lepiej poradził sobie klasyfikator SVM. Dla naiwnego klasyfikatora Bayes'a największym wyzwaniem okazała się grupa 2, znajdująca się w centralnej pozycji, z grupą 1 po lewej i grupą 2 po prawej.

Piąty zestaw, będący prostym przypadkiem dystrybucji Gaussowskiej, nie sprawił problemu wykorzystanym klasyfikatorom – oba uzyskały satysfakcjonujące rezultaty.

W szóstym zestawie, z bardziej nachodzącymi na siebie grupami o dystrybucji Gauss'a, zdecydowanie lepiej sprawdził się naiwny klasyfikator Bayes'a. Różnica jest szczególnie widoczna w przypadku grupy pierwszej i drugiej.

Podsumowanie i wnioski

W ramach zadania projektowego zapoznaliśmy się dogłębnie z metodologią ewaluacji klasyfikatorów. Dowiedzieliśmy się, jakich metryk używa się przy ustalaniu skuteczności klasyfikatora, jak je obliczyć, oraz w jaki sposób wykorzystać zawarte w nich informacje. Wygenerowaliśmy testowe zbiory danych z wykorzystaniem dystrybucji Poisson'a i Cauchy'ego, oraz dokonaliśmy na ich przykładzie ewaluacji naiwnego klasyfikatora Bayes'a oraz klasyfikatora SVM. Odkryliśmy silne oraz słabe punkty testowanych klasyfikatorów. Od strony technicznej zapoznaliśmy się ze środowiskiem RStudio oraz przetwarzaniem zbiorów danych w języku R, jak również z obsługą systemu kontroli wersji Git. Udoskonaliliśmy nasze umiejętności pracy zespołowej, wspólnej realizacji projektów, oraz podziału zadań w grupie.