

## Zad 2

Sporządź w postaci tabeli/tabel i wykresów zestawienie miar klasyfikacji wymienionych powyżej dla wybranego zbioru do klasyfikacji, np. z UCI, Kaggle oraz klasyfikatorów:

- naiwny klasyfikator Bayesa (NaiveBayes)
- SVM (SMO)
- regresji logistycznej (SimpleLogistic)
- 1NN (IBk)
- kNN (IBK - wybrana liczba sąsiadów)
- ...

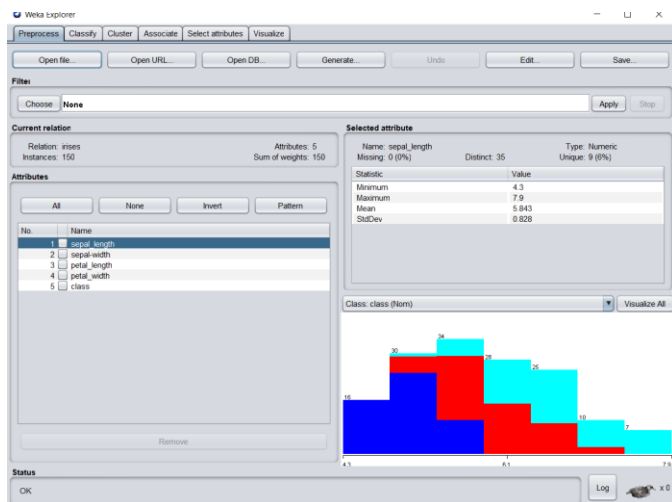
Powtórz zadanie dla kilku rodzajów ewaluacji metod klasyfikacji:

- metoda resubstytucji (uczenie i testowanie na całym zbiorze danych) - tylko do tego raportu
- metoda podziału na część uczącą i testującą (np. 2/3 vs 1/3)
- krosswalidacja (liczba foldsów k między 5 a 10, lub inna)
- metoda leave-one-out (krosswalidacja z k=n, gdzie n to liczba obiektów w całym zbiorze)

## IRIS DATA SET

Zestaw pomiarów kwiatów irysa, udostępniony po raz pierwszy przez Ronalda Fishera w roku 1936. Jeden z najbardziej znanych zbiorów, a zarazem bardzo prosty i użyteczny. Celem jest wytrenowanie systemu, który na podstawie 4 podanych parametrów, poda właściwą klasę kwiatu (jedną z trzech dostępnych).

## Wczytane dane

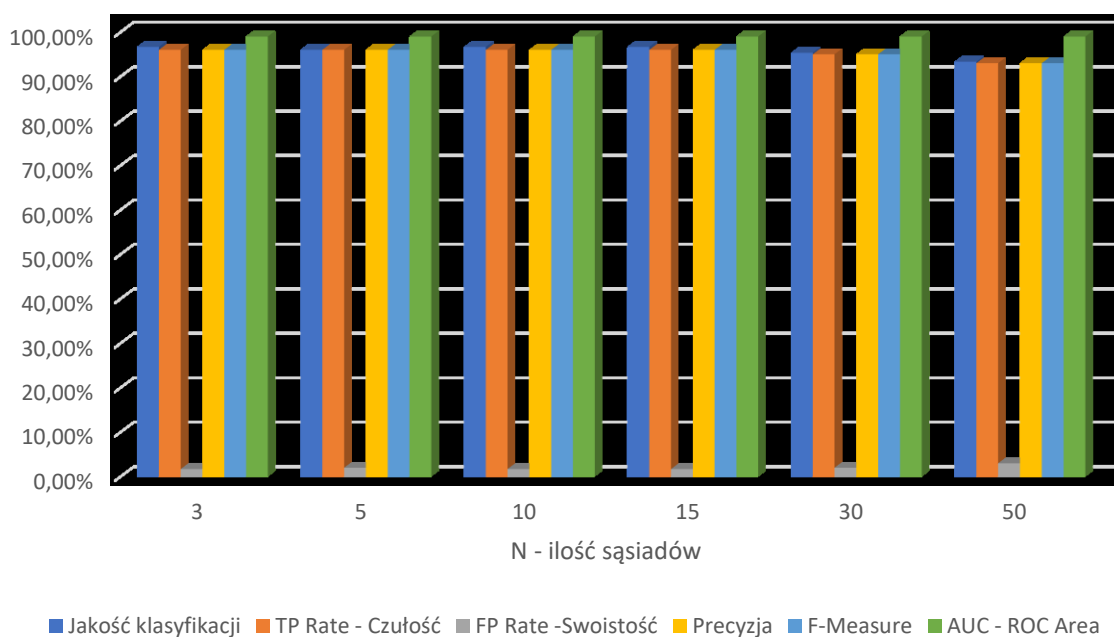


## Klasyfikator IBK

Classifier model (full training set), uwzględniono wartości średnie

N	Jakość klasyfikacji	TP Rate - Czulość	FP Rate - Swoistość	Precyzja	F-Measure	AUC - ROC Area	macierz pomyłek
1	100%	1	0	1	1	1	a b c <-- classified as 50 0 0   a = Iris-setosa 0 50 0   b = Iris-versicolor 0 0 50   c = Iris-virginica
3	96,66%	0,96	0,017	0,96	0,96	0,99	a b c <-- classified as 50 0 0   a = Iris-setosa 0 48 2   b = Iris-versicolor 0 3 47   c = Iris-virginica
5	96%	0,96	0,02	0,96	0,96	0,99	a b c <-- classified as 50 0 0   a = Iris-setosa 0 48 2   b = Iris-versicolor 0 4 46   c = Iris-virginica
10	96,66%	0,96	0,017	0,96	0,96	0,99	a b c <-- classified as 50 0 0   a = Iris-setosa 0 49 1   b = Iris-versicolor 0 4 46   c = Iris-virginica
15	96,66%	0,96	0,017	0,96	0,96	0,99	a b c <-- classified as 50 0 0   a = Iris-setosa 0 49 1   b = Iris-versicolor 0 4 46   c = Iris-virginica
30	95,33%	0,95	0,02	0,95	0,95	0,99	a b c <-- classified as 50 0 0   a = Iris-setosa 0 47 3   b = Iris-versicolor 0 4 46   c = Iris-virginica
50	93,33	0,93	0,03	0,93	0,93	0,99	a b c <-- classified as 50 0 0   a = Iris-setosa 0 48 2   b = Iris-versicolor 0 8 42   c = Iris-virginica

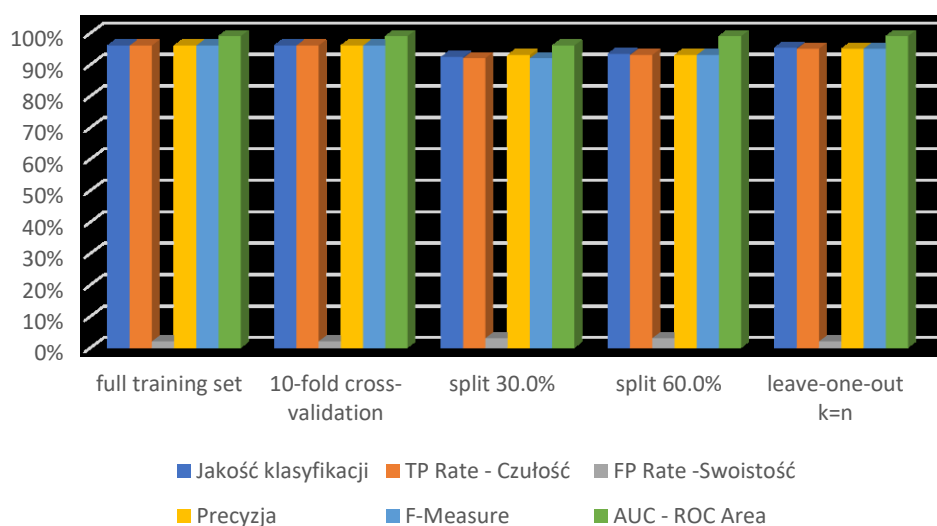
Miary klasyfikacji w zależności od ilości sąsiadów - N



## Klasyfikator Naiwny Bayesa dla wartości średnich

model	Jakość klasyfikacji	TP Rate - Czułość	FP Rate - Swoistość	Precyzja	F- Measure	AUC - ROC Area	macierz pomyłek
<i>full training set</i>	96%	0,96	0,02	0,96	0,96	0,99	a b c <-- classified as 50 0 0   a = Iris-setosa 0 48 2   b = Iris- versicolor 0 4 46   c = Iris-virginica
<i>10-fold cross-validation</i>	96%	0,96	0,02	0,96	0,96	0,99	a b c <-- classified as 50 0 0   a = Iris-setosa 0 48 2   b = Iris- versicolor 0 4 46   c = Iris-virginica
<i>split 30.0%</i>	92,38%	0,92	0,03	0,93	0,92	0,96	a b c <-- classified as 34 0 0   a = Iris-setosa 0 31 1   b = Iris- versicolor 0 7 32   c = Iris-virginica
<i>split 60.0%</i>	93,33%	0,93	0,03	0,93	0,93	0,99	a b c <-- classified as 17 0 0   a = Iris-setosa 0 20 1   b = Iris- versicolor 0 3 19   c = Iris-virginica
<i>leave-one-out k=n</i>	95,33%	0,95	0,02	0,95	0,95	0,99	a b c <-- classified as 50 0 0   a = Iris-setosa 0 47 3   b = Iris- versicolor 0 4 46   c = Iris-virginica

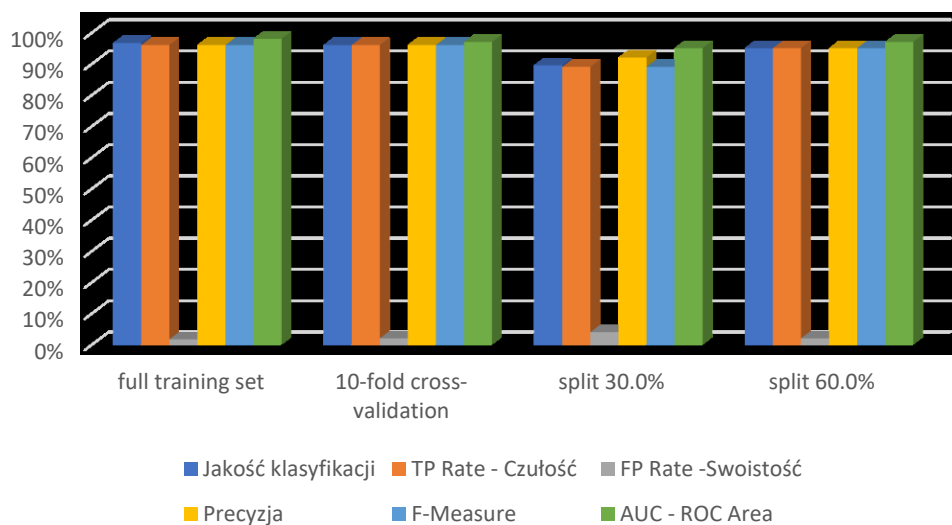
Miary klasyfikacji NaiveBayes



## SVM (SMO)

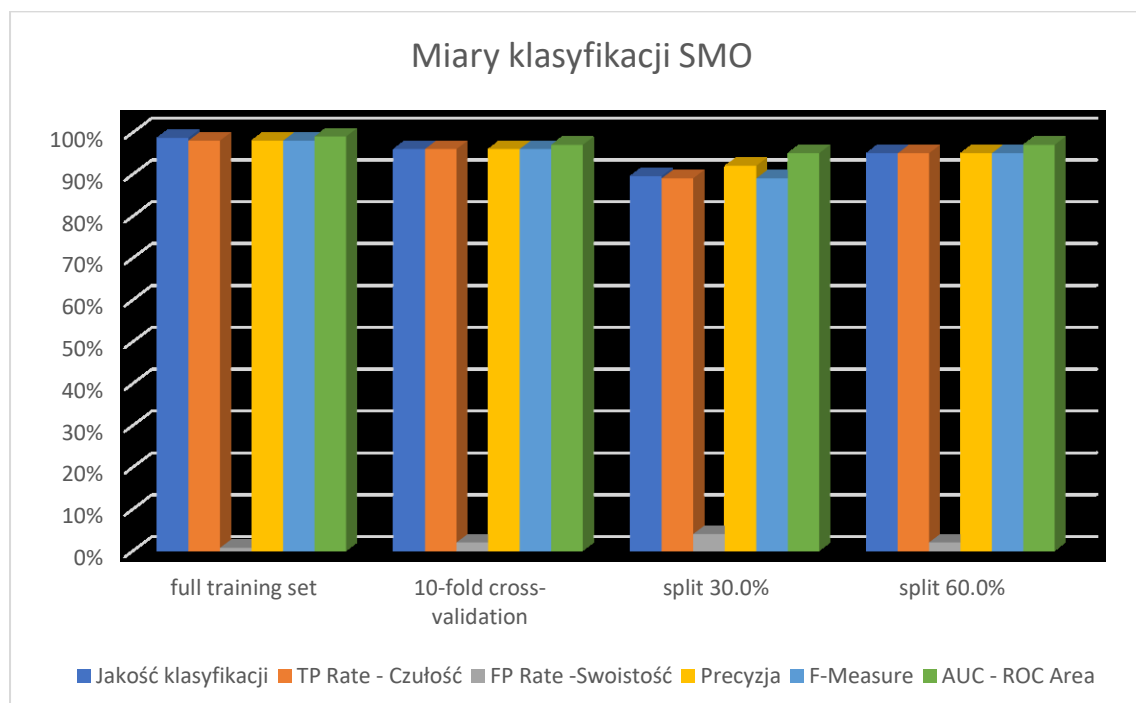
model	Jakość klasyfikacji	TP Rate - Czułość	FP Rate - Swoistość	Precyzja	F- Measure	AUC - ROC Area	macierz pomyłek
<i>full training set</i>	97%	0,96	0,017	0,96	0,96	0,98	a b c <-- classified as 50 0 0   a = Iris-setosa 0 49 1   b = Iris- versicolor 0 4 46   c = Iris-virginica
<i>10-fold cross-validation</i>	96%	0,96	0,02	0,96	0,96	0,97	a b c <-- classified as 50 0 0   a = Iris-setosa 0 49 1   b = Iris- versicolor 0 5 45   c = Iris-virginica
<i>split 30.0%</i>	89,52%	0,89	0,04	0,92	0,89	0,95	a b c <-- classified as 34 0 0   a = Iris-setosa 0 32 0   b = Iris- versicolor 0 11 28   c = Iris-virginica
<i>split 60.0%</i>	95,00%	0,95	0,02	0,95	0,95	0,97	a b c <-- classified as 17 0 0   a = Iris-setosa 0 21 0   b = Iris- versicolor 0 3 19   c = Iris-virginica

Miary klasyfikacji dla SVM (SMO)

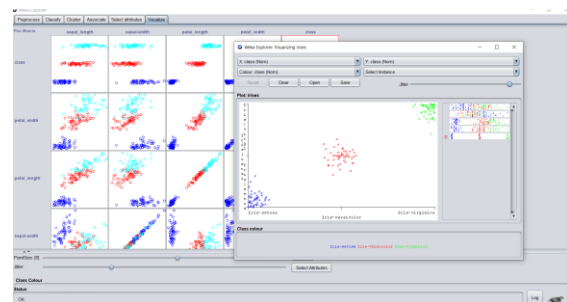
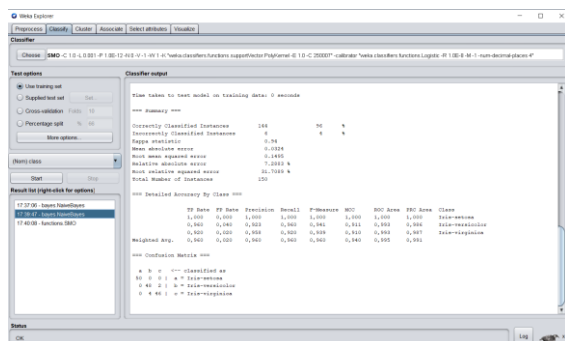


## Regresja logistyczna (SimpleLogistic)

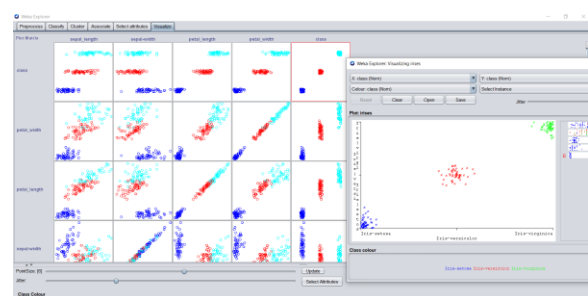
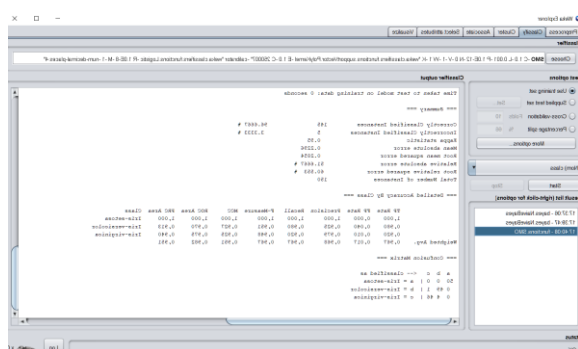
model	Jakość klasyfikacji	TP Rate - Czułość	FP Rate - Swoistość	Precyzja	F- Measure	AUC - ROC Area	macierz pomyłek
<i>full training set</i>	99%	0,98	0,007	0,98	0,98	0,99	a b c <-- classified as 50 0 0   a = Iris-setosa 0 49 1   b = Iris- versicolor 0 1 49   c = Iris-virginica
<i>10-fold cross-validation</i>	96%	0,96	0,02	0,96	0,96	0,97	a b c <-- classified as 50 0 0   a = Iris-setosa 0 49 1   b = Iris- versicolor 0 5 45   c = Iris-virginica
<i>split 30.0%</i>	89,52%	0,89	0,04	0,92	0,89	0,95	a b c <-- classified as 34 0 0   a = Iris-setosa 0 32 0   b = Iris- versicolor 0 11 28   c = Iris-virginica
<i>split 60.0%</i>	95,00%	0,95	0,02	0,95	0,95	0,97	a b c <-- classified as 17 0 0   a = Iris-setosa 0 21 0   b = Iris- versicolor 0 3 19   c = Iris-virginica



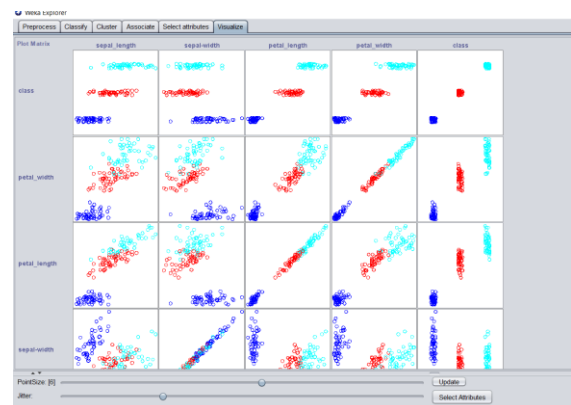
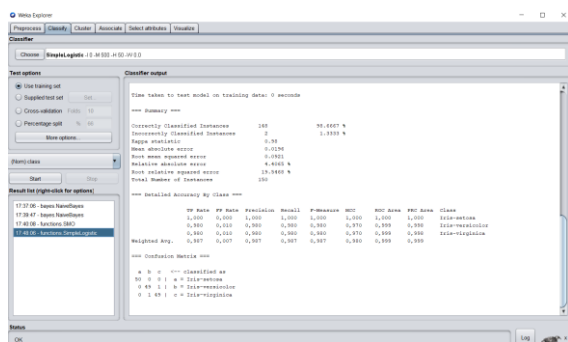
## Dodatkowe screeny



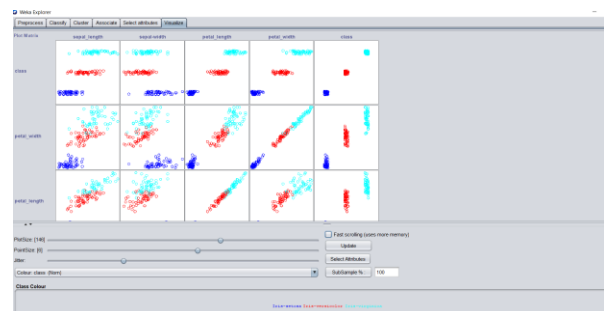
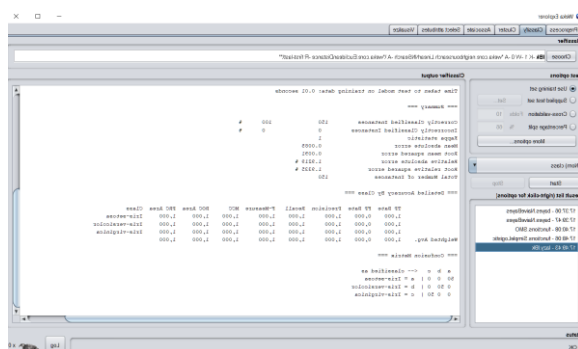
## Klasyfikator SVM (SMO)



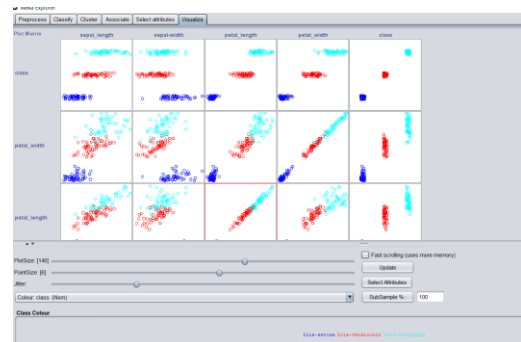
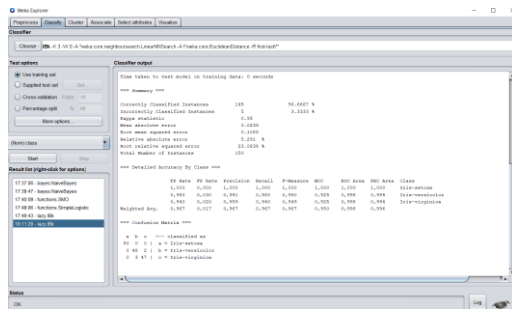
## Klasyfikator regresji logistycznej (SimpleLogistic)



## 1NN (IBk)



## IBk, KNN = 3



## IBk, KNN = 5

