KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

dr inż. Artur Klepaczko

Eksploracja danych
Ocena klasyfikacji

Zadanie nr 13 – Studia podyplomowe „Przetwarzanie i analiza obrazów biomedycznych”

Prezentacja multimedialna
współfinansowana przez Unię Europejską
w ramach Europejskiego Funduszu Społecznego
w projekcie

„Innowacyjna dydaktyka bez ograniczeń – zintegrowany rozwój Politechniki Łódzkiej – zarządzanie Uczelnią, nowoczesna oferta edukacyjna i wzmacniania zdolności do zatrudniania osób niepełnosprawnych”

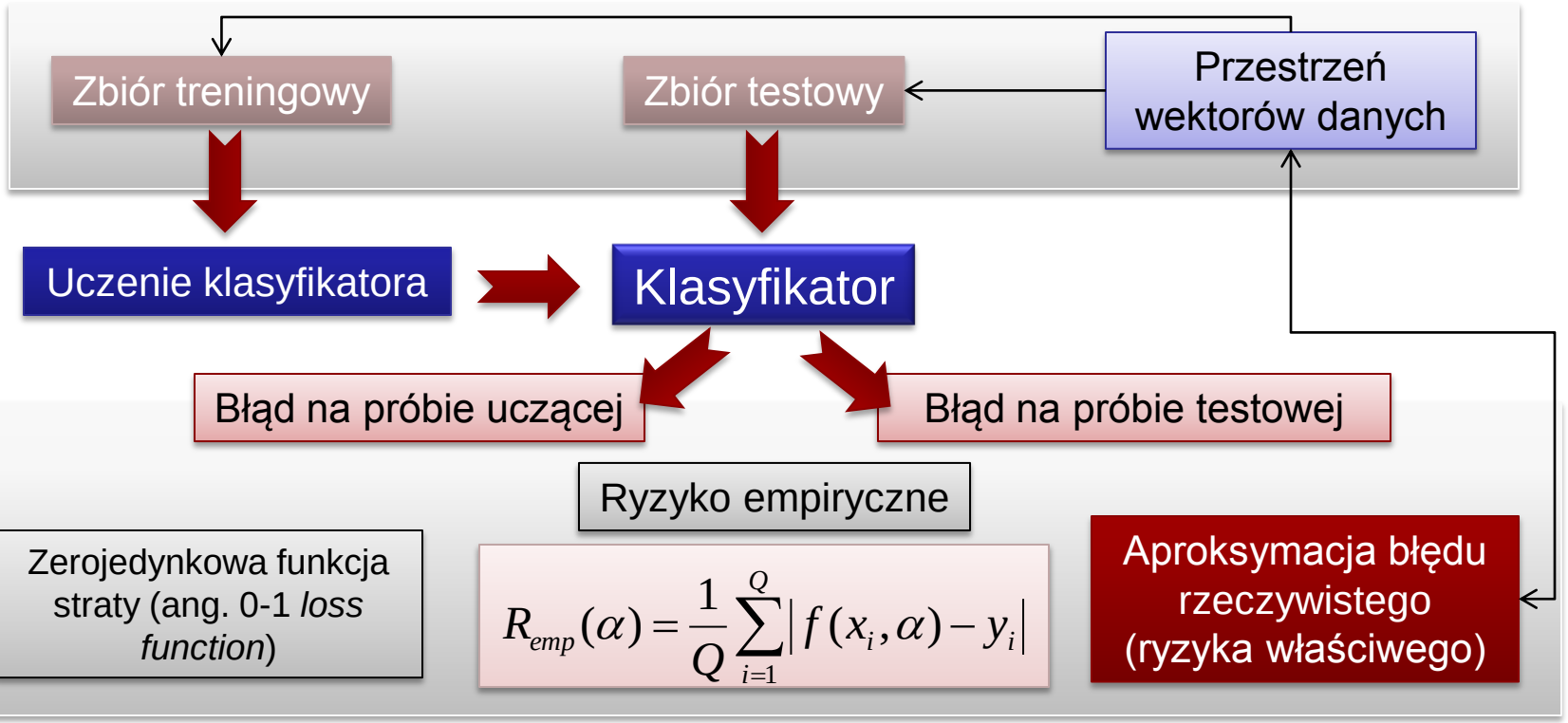
Politechnika Łódzka
Instytut Elektroniki

90-924 Łódź, ul. Żeromskiego 116,
tel. 042 631 28 83
www.kapitalludzki.p.lodz.pl

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

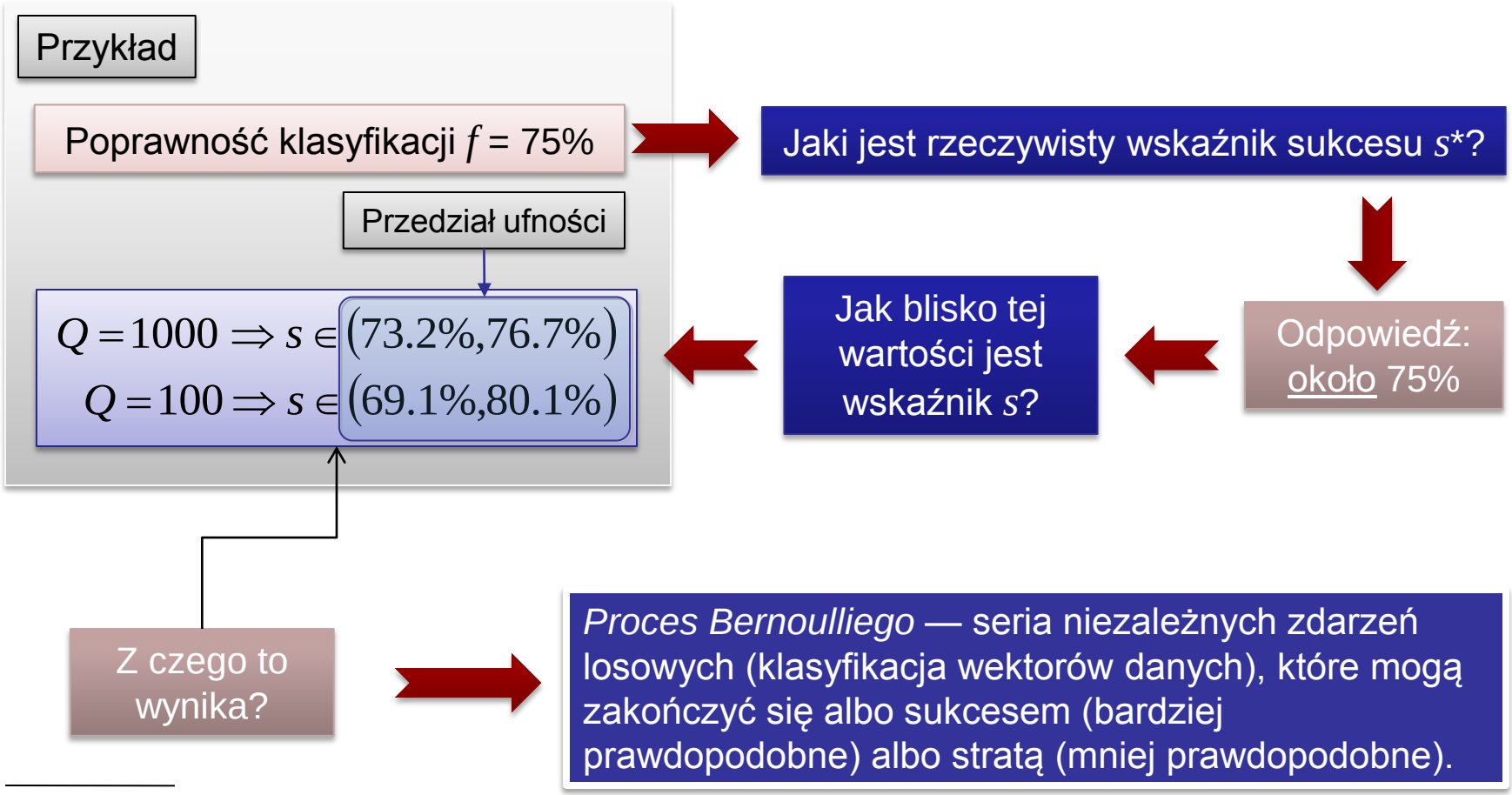
Ocena algorytmu klasyfikacji

Na ile algorytm nauczony na podstawie próby uczącej będzie zdolny do prawidłowej predykcji klasy nowych wektorów danych?



Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Przedział ufności



*Wskaźnik sukcesu = 1 – błąd klasyfikacji

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

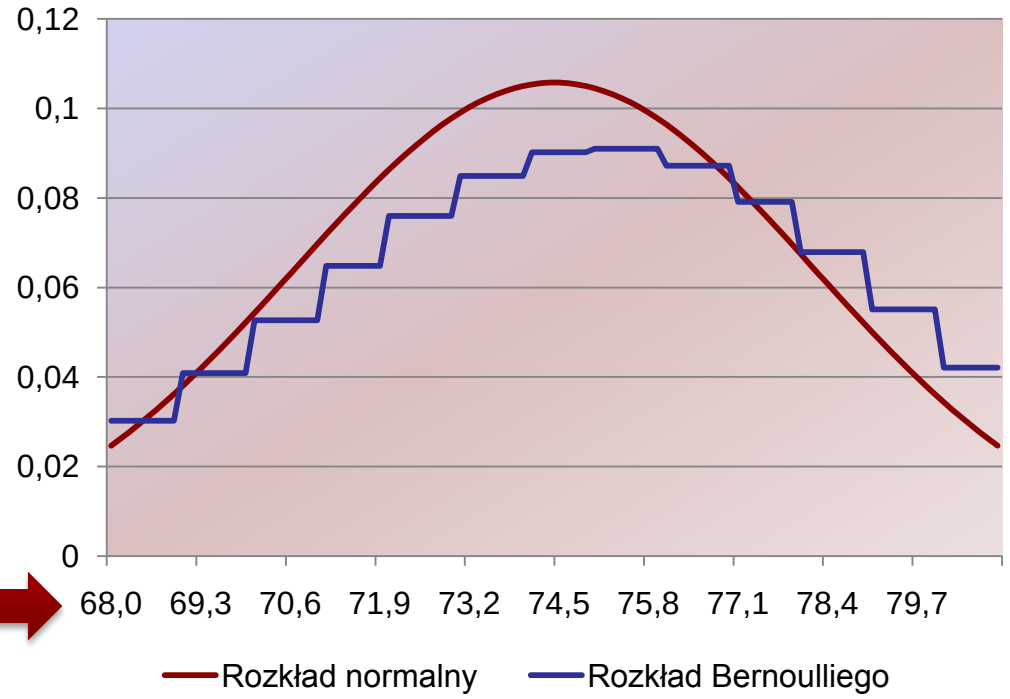
Proces Bernoulliego a rozkład normalny

Średnia szansa na to, aby pojedyncza próba należąca do procesu Bernoulliego zakończyła się sukcesem wynosi s.

Średnia liczba sukcesów dla serii prób Bernoulliego wynosi również s.

W przypadku dużej serii zdarzeń, proces Bernoulliego zbliża się do rozkładu Gaussa.

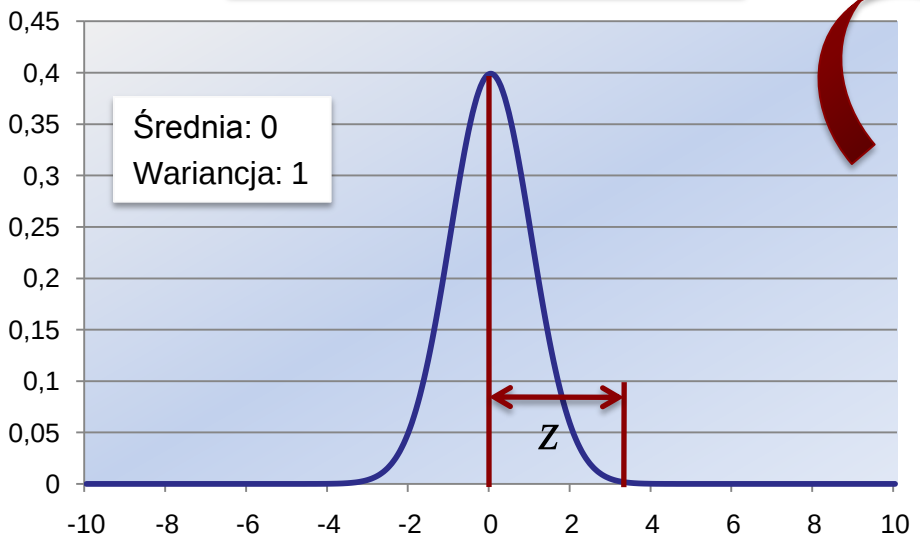
Średnia poprawność klasyfikacji [%] dla różnych serii zdarzeń



Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Obliczanie przedziału ufności

Rozkład normalny $N(0,1)$



Granice przedziałów ufności

$\Pr[X \geq z]$	z
0,1%	3,09
0,5%	2,58
1%	2,33
5%	1,65
10%	1,28
20%	0,84

Wynik klasyfikacji

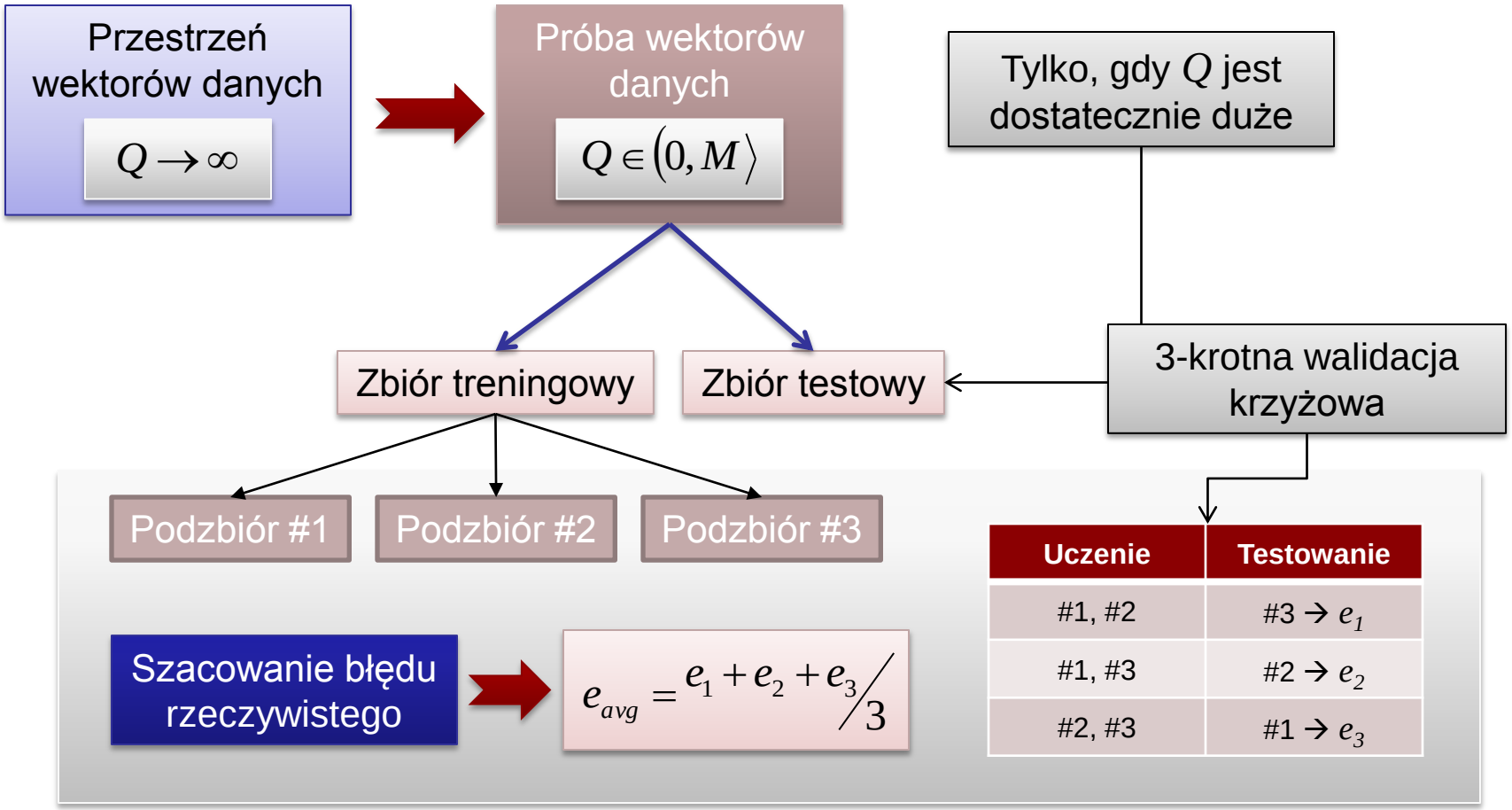
$$s = \left(f + \frac{z^2}{2Q} \pm z\sqrt{\frac{f}{Q} - \frac{f^2}{Q} + \frac{z^2}{4Q^2}} \right) \Big/ \left(1 + \frac{z^2}{Q} \right)$$

Z tabeli

Prawdopodobieństwo tego, że zmienna X przyjmie wartość oddaloną o 2,33 od średniej (powyżej lub poniżej) wynosi 2%.

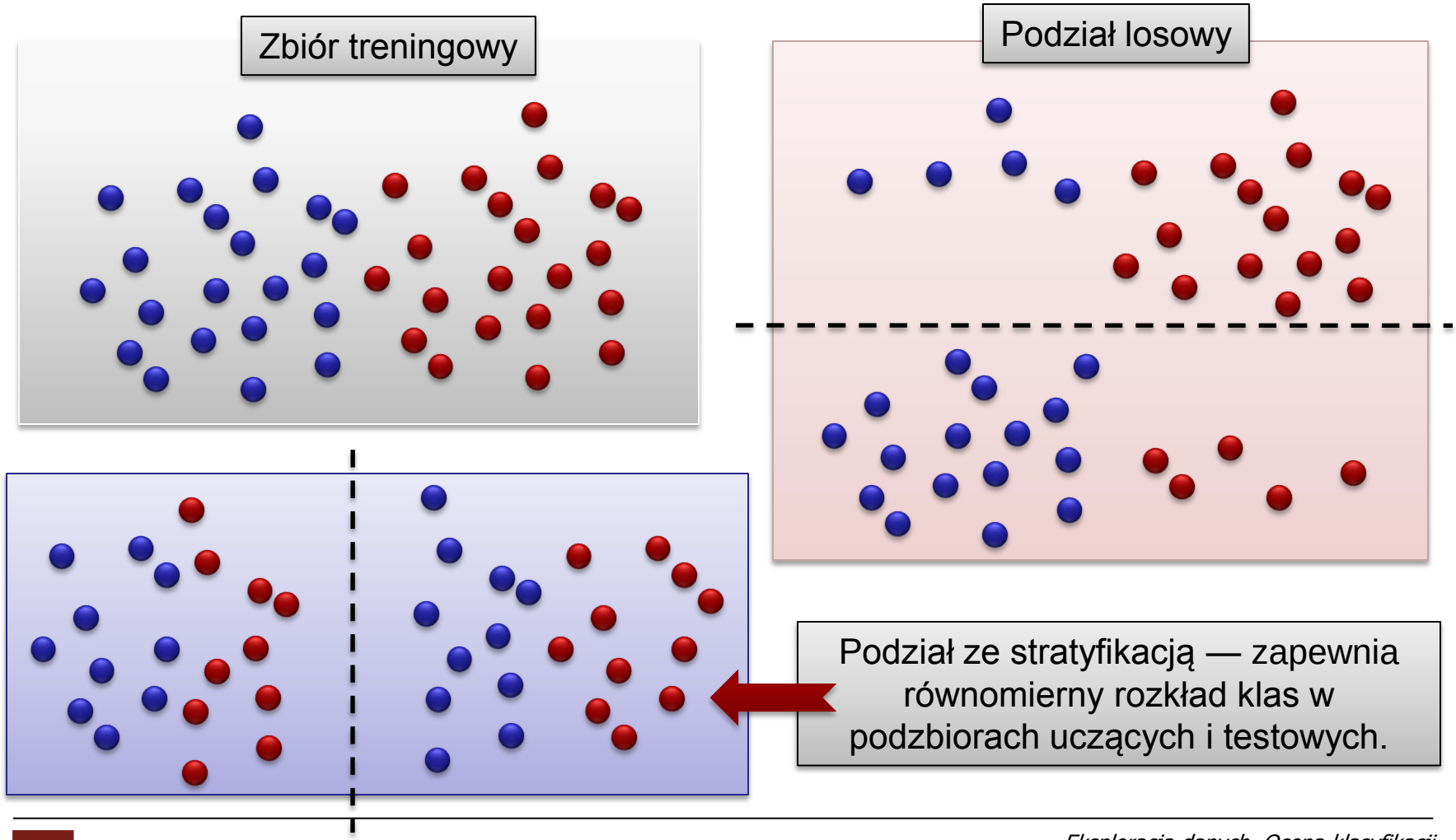
Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Próba losowa wektorów danych



Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Walidacja krzyżowa ze stratyfikacją



Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Szczególny przypadek: *Leave-one-out*

Metoda leave-one-out: Q –krotna walidacja krzyżowa, gdzie Q jest liczbą wszystkich wektorów danych.

Zalety

- Największa możliwa część zbioru danych jest dostępna w fazie uczenia.
- Oszacowany błąd zawsze taki sam dla danego zbioru danych (podział zbioru treningowego jest deterministyczny).

Wady

- Duża złożoność obliczeniowa, szczególnie w przypadku większych zbiorów danych i skomplikowanych algorytmów uczenia.
- Nierównomierne rozłożenie klas w zbiorach testowych i treningowych.

Metoda *bootstrap 0.632*

Losowanie Q wektorów ze zwracaniem

Zbiór Q wektorów danych

Wektory treningowe (mogą się powtarzać)

Prawdopodobieństwo wylosowania wektora w pojedynczej próbie:

$1/Q$

Wektory testowe (niewylosowane)

Prawdopodobieństwo niewylosowania wektora w pojedynczej próbie:

$1-1/Q$

Prawdopodobieństwo niewylosowania wektora w żadnej z Q prób:

$$\left(1-1/Q\right)^{Q} = e^{-1} = 0.368$$

Podstawa logarytmu naturalnego

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Szacowanie błędu rzeczywistego w metodzie *bootstrap*

Dla dostatecznie dużego Q zbiór treningowy zawiera ok. 63,2% wektorów z oryginalnego zbioru danych (przy 90% w przypadku zastosowania 10-krotnej walidacji krzyżowej).

Błąd klasyfikacji popełniony na zbiorze uczącym: e_{train}

Błąd klasyfikacji popełniony na zbiorze testowym: e_{test}

Oszacowanie zbyt optymistyczne

Oszacowanie zbyt pesymistyczne

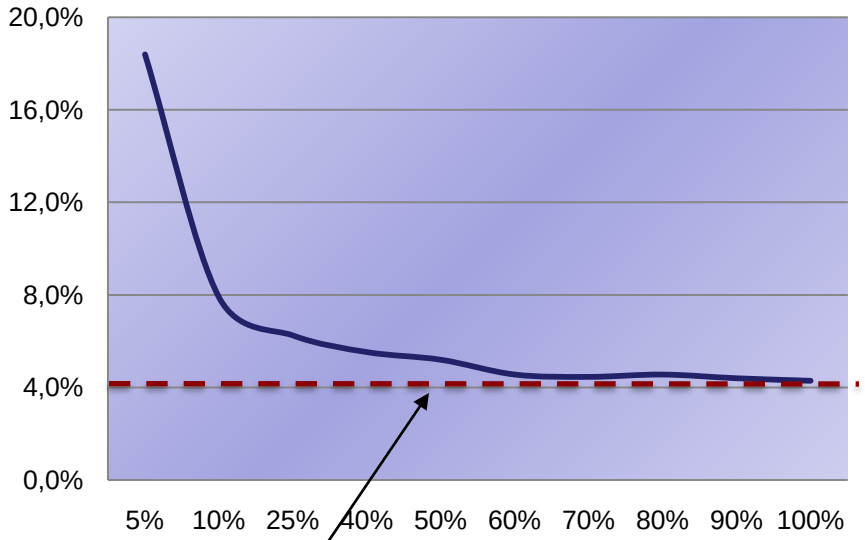
$$e_{avg} = 0.632 \cdot e_{test} + 0.368 \cdot e_{train}$$

W przypadku nierównomiernego rozłożenia klas w zbiorach testowym i treningowym oszacowanie błędu rzeczywistego nadal może być zbyt pesymistyczne.

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Krzywa uczenia

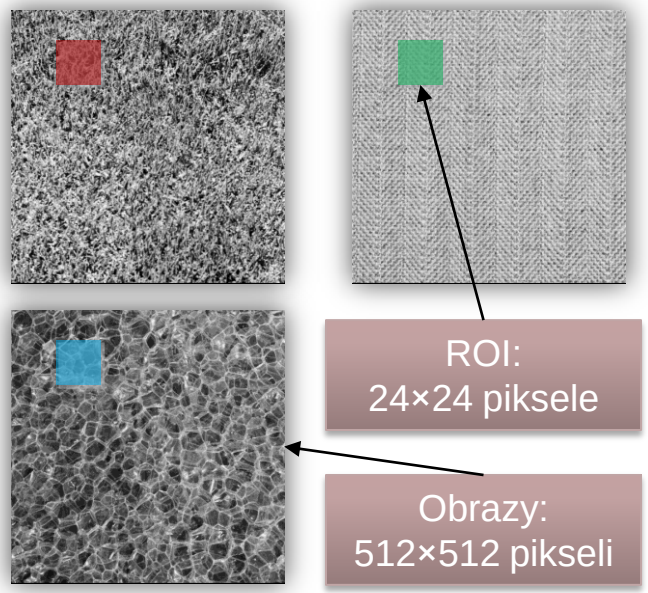
Błąd klasyfikacji w zależności od liczby wektorów



Prawdopodobna wartość błędu rzeczywistego

Procent całkowitej liczby wektorów biorących udział w uczeniu i testowaniu

Przykład



ROI:
24×24 piksele

Obrazy:
512×512 pikseli

Zbiór danych:
- 3 klasy
- 256 wektorów danych w klasie

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Wybór klasyfikatora

	Zbiory obrazów tekstur (24×24 piksele, 128 wektorów w klasie)		
	grass, weave, plastic	*bark, brick, rafia*	*brick, sand, clothe*
Support vector machine (liniowa funkcja jądra)	**5,9%**	**2,4%**	**2,3%**
Regresja logistyczna	**4,2%**	**1,6%**	**3,9%**
Algorytm 1-R	**21,6%**	**6,3%**	**18,0%**

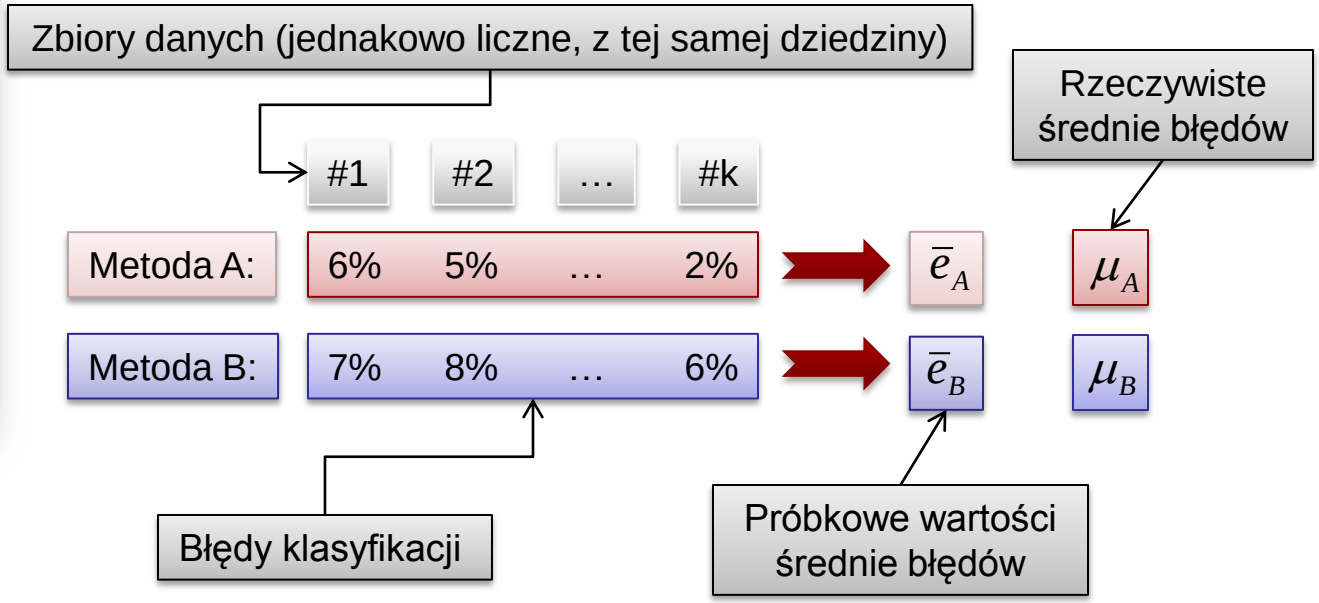
Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Statystyka *t*–Studenta

Hipoteza zerowa: różnica między uzyskanymi wskaźnikami poprawności klasyfikacji jest przypadkowa; w rzeczywistości wszystkie testowane algorytmy są jednakowo dokładne.

Wiliam Seally Gosset, ps. Student, 1876–1937

Statystka *t*–Studenta pozwala zweryfikować słuszność tej hipotezy.



Sparowany test Studenta

W przypadku gdy rozmiary próbek są duże, ich średnie mają rozkład normalny.

…ale rozmiary nie są duże ☹ → Średnie mają rozkład Studenta.

1. Jeśli poszczególne wyniki dla metod A oraz B uzyskano dla tych samych zbiorów danych, to stosujemy sparowany test Studenta.
2. Analizujemy zmienną:

$$\bar{d} = \bar{e}_A - \bar{e}_B$$

3. Jeżeli hipoteza zerowa jest słuszna, to:

$$\mu_A = \mu_B$$

Standaryzacja zmiennej losowej dla rozkładu Studenta

$$t = \frac{\bar{e}_A - \mu_A}{\sqrt{\sigma_{e_A}^2 / k}}$$

$$t = \frac{\bar{d}}{\sqrt{\sigma_d^2 / k}}$$

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Przedziały ufności dla zmiennej *t*–Studenta

Granice przedziałów ufności

Wartości dla przedziałów jednostronnych

Tabela dla k=10 (9 stopni swobody)

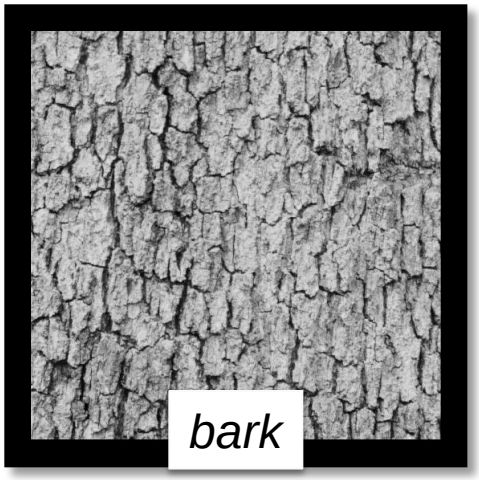
Rozkład Studenta	
$\Pr[X \geq z]$	z
0,1%	4,30
0,5%	3,25
1%	2,82
5%	1,83
10%	1,38
20%	0,88

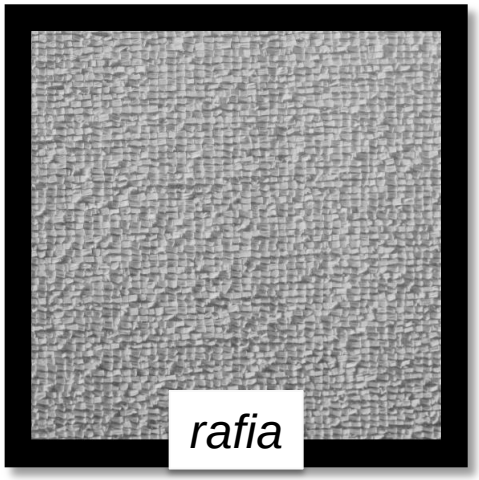
Rozkład Guassa	
$\Pr[X \geq z]$	z
0,1%	3,09
0,5%	2,58
1%	2,33
5%	1,65
10%	1,28
20%	0,84

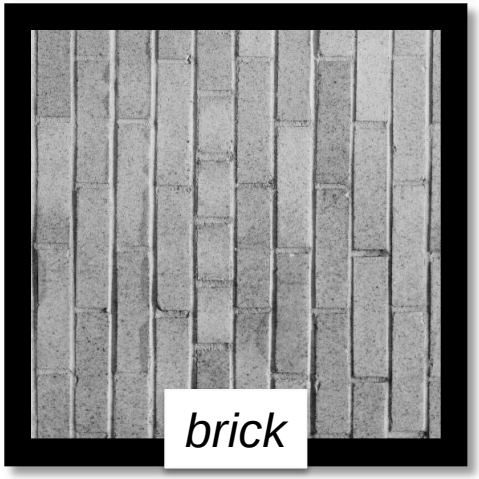
1. Decydujemy się na jakiś poziom ufności → 99%.
2. Obliczamy wartość zmiennej t → t_{exp}.
3. Jeżeli $t > z$ lub $t < -z$ to odrzucamy hipotezę zerową.

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Przykład — zbiór obrazów tekstur







ROI: 24×24 piksele

Liczba wektorów w klasie: 128

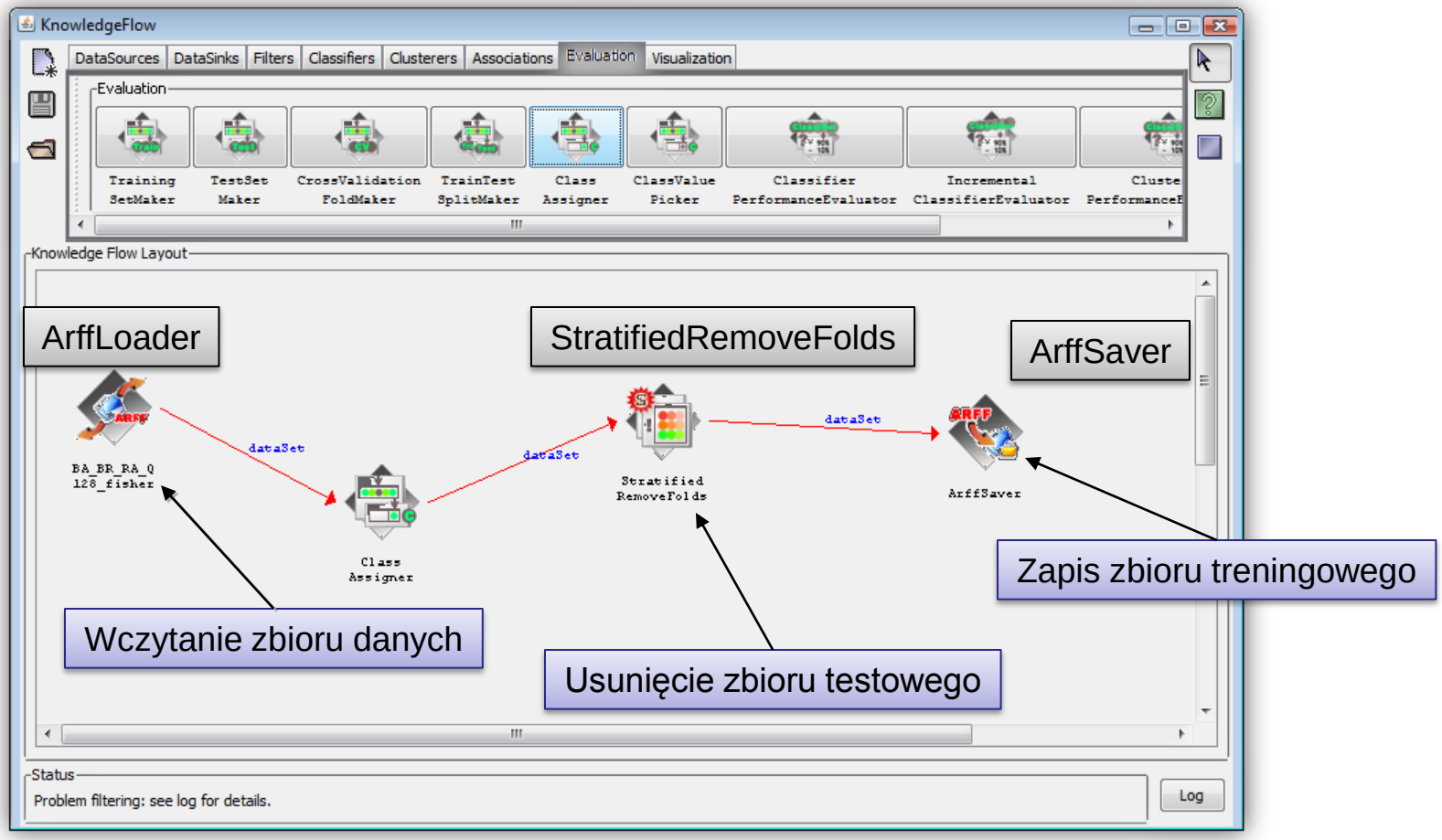
Liczba cech tekstury:10 (selekcja przy użyciu współczynnika Fishera)

Przebieg eksperymentu:

1. Utworzenie dziesięciu par zbiorów treningowych oraz testowych (Weka)
2. Uczenie i testowanie klasyfikatora typu SVM (Weka) → zanotowanie wyników (MS Excel)
3. Uczenie i testowanie klasyfikatora regresyjno–logistycznego (Weka) → zanotowanie wyników (MS Excel)
4. Porównanie efektywności klasyfikatorów przy użyciu testu Studenta (MS Excel)

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

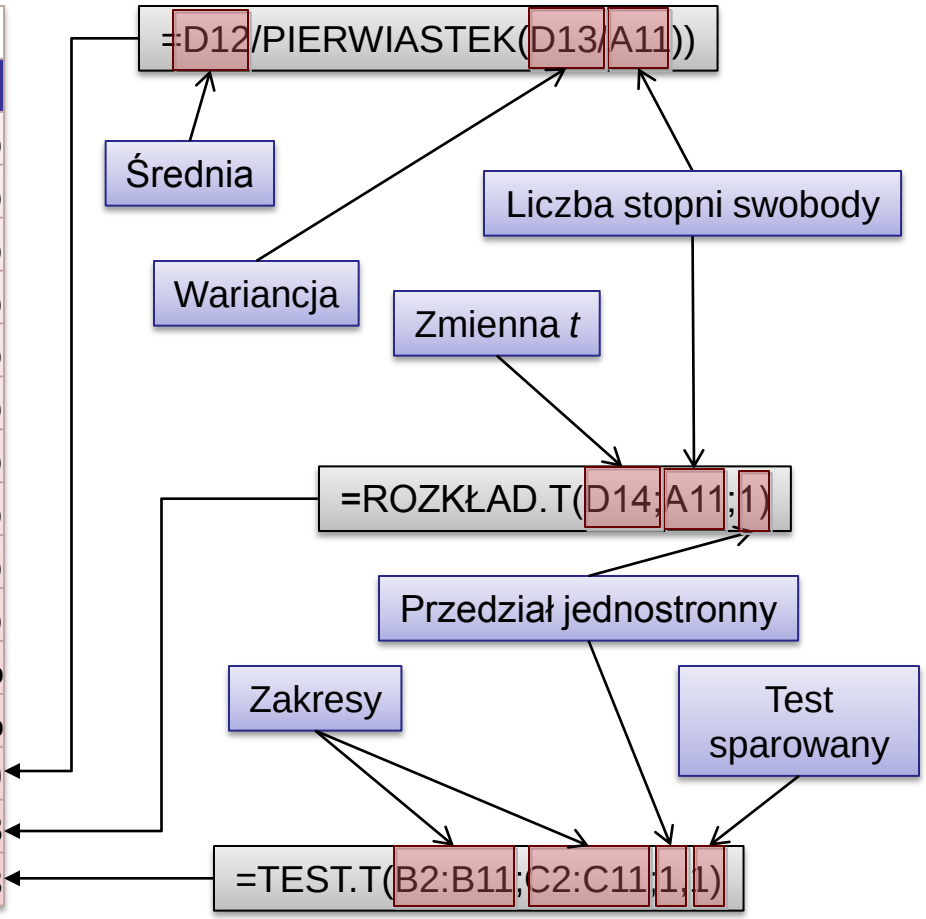
KnowledgeFlow — przygotowanie eksperymentu



Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Wyniki doświadczenia (obliczenia w MS Excel)

	A	B	C	D
1	Lp.	SVM	Logistic	SVM - Logistic
2	1	2,56%	5,13%	-2,56%
3	2	0,00%	0,00%	0,00%
4	3	0,00%	0,00%	0,00%
5	4	2,56%	2,56%	0,00%
6	5	2,63%	0,00%	2,63%
7	6	2,63%	2,63%	0,00%
8	7	2,63%	5,26%	-2,63%
9	8	0,00%	0,00%	0,00%
10	9	2,63%	0,00%	2,63%
11	10	5,26%	0,00%	5,26%
12			**Średnia**	**0,53%**
13			**Wariancja**	**0,06%**
14			**t**	**0,70**
15			**Rozkład.T**	**0,249978**
16			**Test.T**	**0,250843**



Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Analiza wyników doświadczenia

Rozkład Studenta

Pr[$X \geq z$]	z
0,1%	4,30
0,5%	3,25
1%	2,82
5%	1,83
10%	1,38

Przy poziomie ufności równym 99% (przedział dwustronny), aby odrzucić hipotezę o równej skuteczności porównywanych klasyfikatorów, musiałby być spełniony warunek, iż $t > 3,25$.

	A	B	C	D
1	Lp.	SVM	Logistic	SVM - Logistic
2	1	2,56%	5,13%	-2,56%
3	2	0,00%	0,00%	0,00%
4	3	0,00%	0,00%	0,00%
5	4	2,56%	2,56%	0,00%
6	5	2,63%	0,00%	2,63%
7	6	2,63%	2,63%	0,00%
11	10	5,26%	0,00%	5,26%
12			Średnia	0,53%
13			Wariancja	0,06%
			t	**0,70**
			Rozkład.T	**0,249978**
			Test.T	**0,250843**

Istnieje 25% prawdopodobieństwo, że zmienna Studenta przyjmie wartość równą lub większą od 0,7.

Wniosek: metoda SVM, jak i regresja logistyczna są jednakowo dobre do klasyfikacji zbioru obrazów tekstur *bark, brick, rafia*.

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Szacowanie prawdopodobieństwa

Nr wektora	Klasyfikator probabilistyczny		Klasyfikator dyskryminacyjny
	Klasa A	Klasa B	
1	95%	5%	A
2	92%	8%	A
3	20%	80%	B
…	…	…	…
45	51%	49%	A
46	49%	51%	B

Niepewność wnioskowania o klasie wektorów. Wektory 45 i 46 mogą z niemal równym prawdopodobieństwem należeć do A albo do B.

Jednoznaczne wskazanie na klasę, bez żadnej sugestii dotyczącej potencjalnej pomyłki.

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Kwadratowa funkcja straty

W przypadku klasyfikatorów „nieostrych”, wyznaczających prawdopodobieństwo przynależności wektorów do poszczególnych klas, stosuje się inne niż zerojedynkowe funkcje straty.

Prawdziwa klasa wektora

$$[a_1, a_2, \ldots, a_K]$$

$$\mathbf{x}_i \in C_c \Rightarrow a_c = 1, a_{j \neq c} = 0$$

Wyznaczone prawdopodobieństwa przynależności do klasy

$$[p_1, p_2, \ldots, p_K]$$

$$\sum_{j=1}^{K} p_j = 1$$

$$l_{kw} = \frac{1}{Q} \sum_{i=1}^{Q} \sum_{j=1}^{K} \left(p_j - a_j \right)^2$$

Minimalizacja kwadratowej funkcji straty ma znaczenie głównie w przypadku, gdy prawdziwa klasa wektorów danych również pochodzi z pewnego rozkładu prawdopodobieństwa.

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Koszty błędów

Podczas uczenia klasyfikatorów najczęściej nie jest brana pod uwagę różny rodzaj popełnianych błędów i związany z nimi koszt.

Przykład

system decyzyjny na podstawie informacji o pacjentach (dane metrykalne, wyniki badań, objawy) klasyfikuje ich do jednej z dwu klas: **chory** / **zdrowy**.

Błąd polegający na uznaniu chorego pacjenta za zdrowego może być bardziej kosztowny niż decyzja, iż zdrowy pacjent jest chory.

Inne przykłady:

- decyzja o przyznaniu pożyczki,
- wykrywanie uszkodzeń,
- identyfikacja samolotów nieprzyjaciela.

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Możliwe wyniki predykcji klasy

Przypadek 2 klas

Rzeczywista klasa \ Wynik klasyfikacji	Tak (1)	Nie (0)
Tak (1)	**TP**	**FN**
Nie (0)	**FP**	**TN**

Oznaczenie danej klasy wektorów etykietą TAK lub NIE zależy od kontekstu.

Wielkość błędu klasyfikacji

$$e = 1 - \frac{TP + TN}{TP + TN + FP + FN}$$

TP — *true positive*

FP — *false positive*

TN — *true negative*

FN — *false negative*

Iloraz *TP*

$$\frac{TP}{TP + FN}$$

Iloraz *FP*

$$\frac{FP}{FP + TN}$$

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Macierz pomyłek

Klasyfikator 1-R

Wynik klasyfikacji

Rzeczywista klasa

	A	B	C	Razem
A	113	0	15	128
B	3	125	0	128
B	6	0	122	128
Razem	122	125	137	384

Wyniki uzyskane dla zbioru obrazów tekstur (*bark, brick, rafia*)

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         360               93.75 %
Incorrectly Classified Instances        24                6.25 %
Kappa statistic                          0.9063
```

Na ile tak naprawdę klasyfikator jest skuteczny?

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Statystyka Kappa

Klasyfikator losowy

Rzeczywista klasa \ Wynik klasyfikacji	A	B	C	Razem
A	45	58	25	128
B	36	37	55	128
B	67	15	46	128
Razem	148	110	126	384

Liczba poprawnych predykcji klasyfikatora 1-*R*

Liczba poprawnych wskazań klasyfikatora losowego

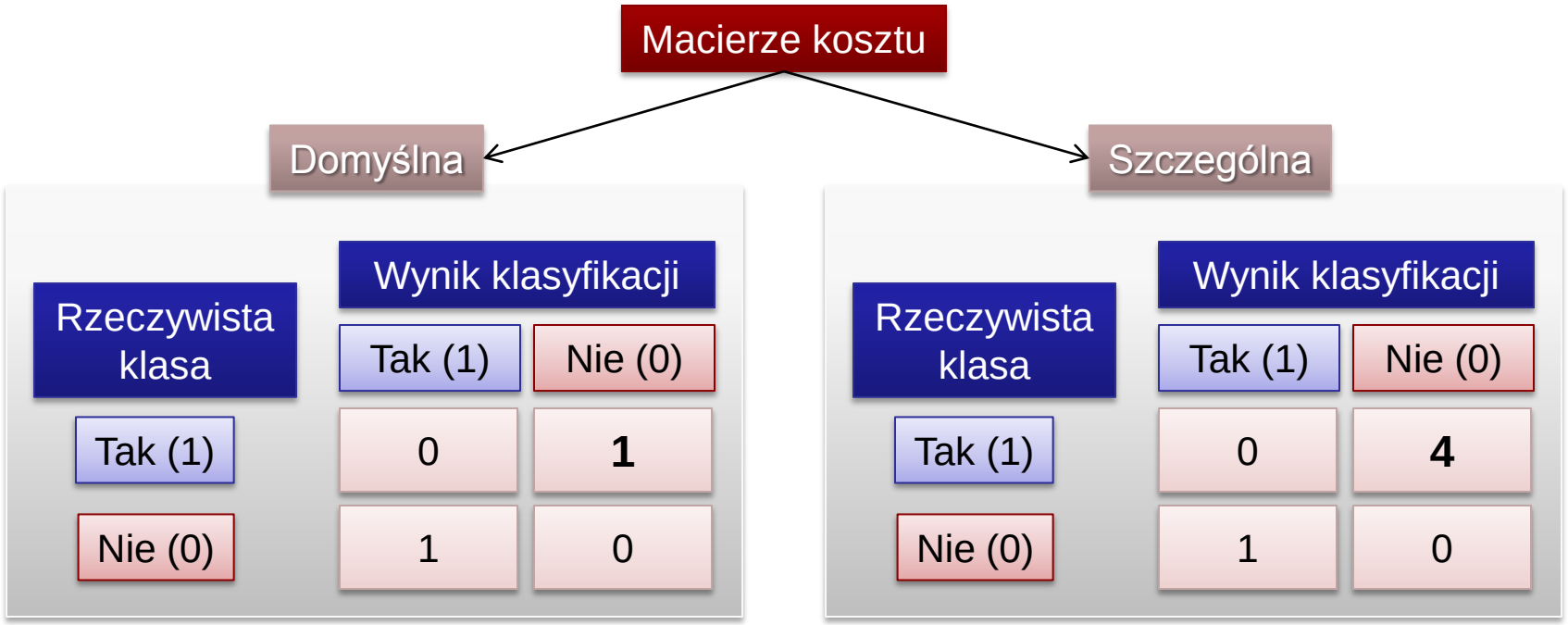
Liczność zbioru danych

$$\kappa = \frac{L_{1-R} - L_{losowy}}{Q - L_{losowy}}$$

$L_{1-R} = 360$ $L_{losowy} = 128$

$Q = 384$ $\kappa = 0{,}9063$

Ocena wyniku uwzględniająca koszt



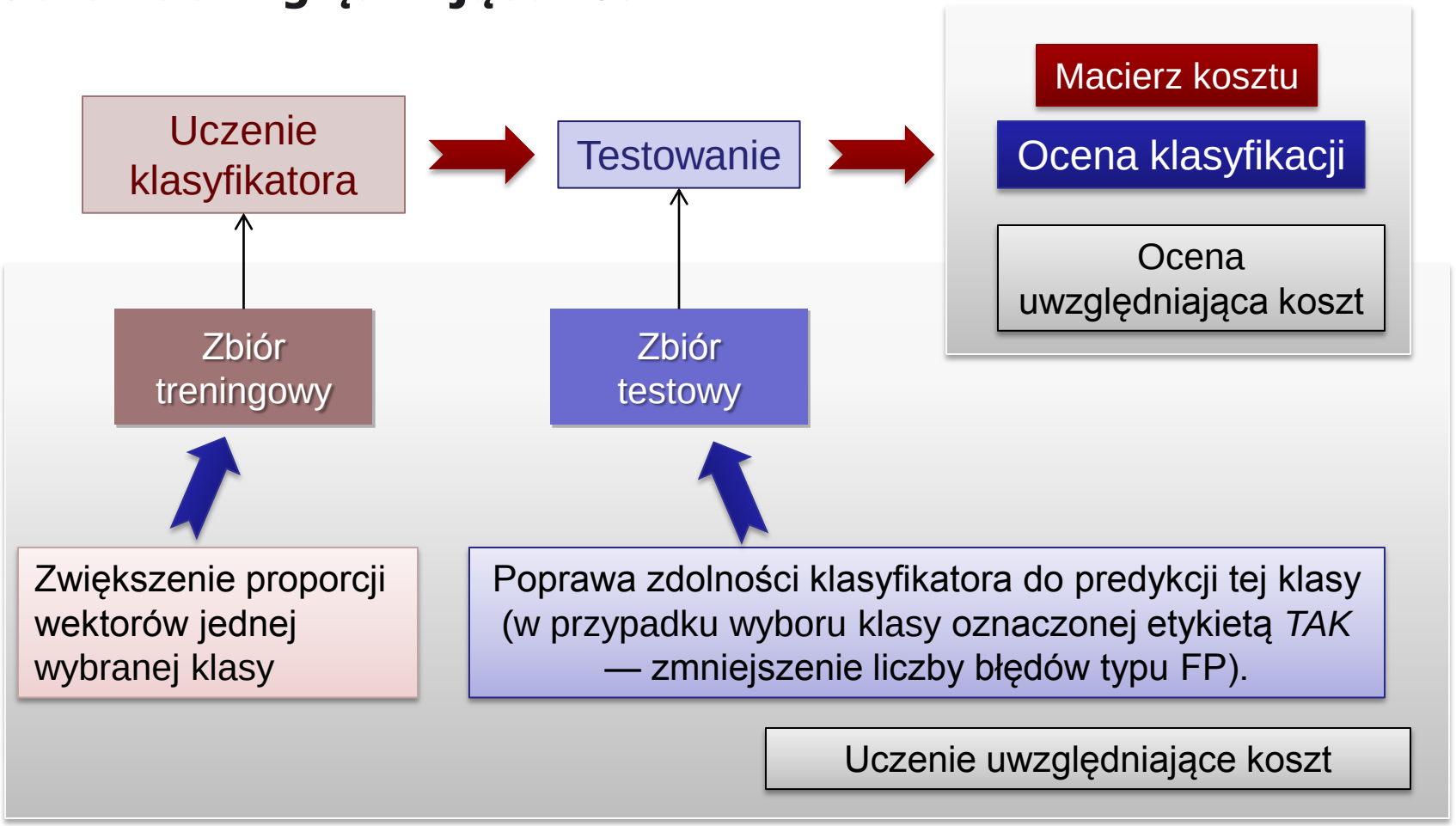
Rzeczywista klasa	Wynik klasyfikacji: Tak (1)	Wynik klasyfikacji: Nie (0)
Tak (1)	0	**1**
Nie (0)	1	0

Rzeczywista klasa	Wynik klasyfikacji: Tak (1)	Wynik klasyfikacji: Nie (0)
Tak (1)	0	**4**
Nie (0)	1	0

- Przy ocenie klasyfikatora brany pod uwagę jest koszt błędów, nie zaś sam wskaźnik sukcesu.
- Model kosztu może uwzględniać także koszt użycia systemu uczącego się, czy też pozyskania danych.
- W przypadku klasyfikatorów probabilistycznych, wektory testowe przypisuje się do najbardziej prawdopodobnej klasy przy jednoczesnym zapewnieniu minimum kosztu decyzji klasyfikującej.

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Uczenie uwzględniające koszt



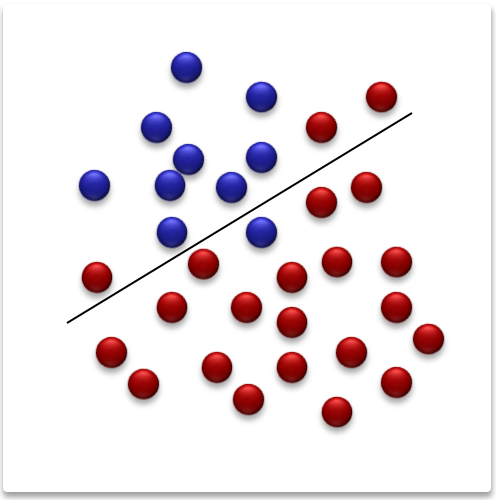
Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Zwiększanie ilorazu TP

Przestrzeń cech

Pomiar dokładny (duży koszt)	Pomiar zgrubny (mały koszt)
$\left[x_1^1, x_1^2, x_1^3, \ldots\right.$	$\left.\ldots\ x_1^{N-2}, x_1^{N-1}, x_1^N\right]$
$\left[x_i^1, x_i^2, x_i^3, \ldots\right.$	$\left.\ldots\ x_i^{N-2}, x_i^{N-1}, x_i^N\right]$
$\left[x_Q^1, x_Q^2, x_Q^3, \ldots\right.$	$\left.\ldots\ x_Q^{N-2}, x_Q^{N-1}, x_Q^N\right]$

Klasyfikator



Liczba wektorów w zbiorze danych (Q) = 10^6

Klasa A (*TAK*) — 1000 wektorów danych

Klasa B (*NIE*) — 999000 wektorów danych

Zadaniem klasyfikatora jest wyznaczenie podzbioru wektorów danych, w którym iloraz TP będzie większy.

Wydobywanie informacji (ang. *information retrieval*, IR)

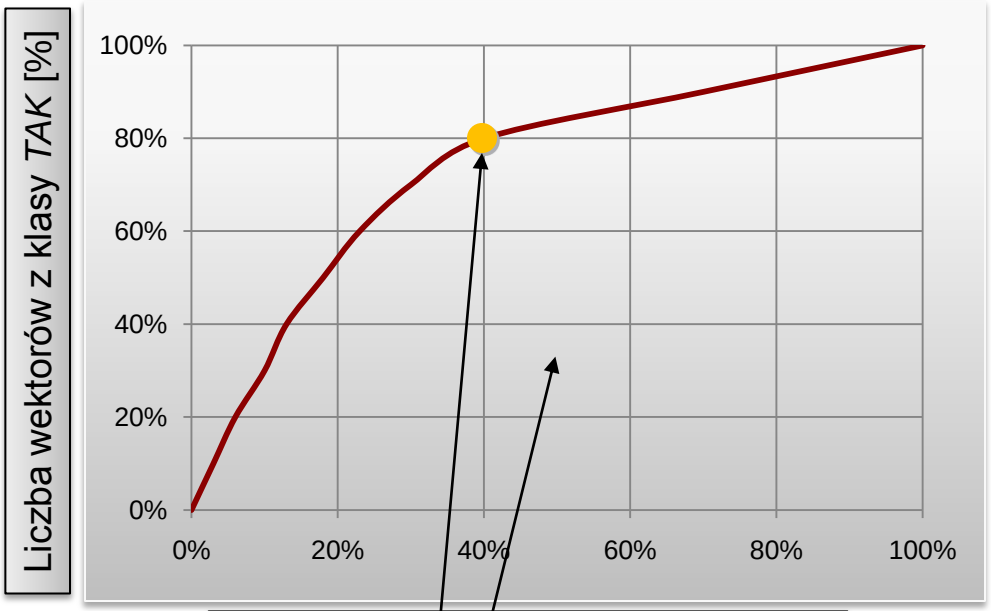
Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Wykres wzrostu

Naiwny klasyfikator Bayesa

Pozycja w rankingu	Wynik klasyfikacji	Rzeczywista klasa
1	0,99	TAK
2	0,98	TAK
3	0,97	TAK
4	0,95	NIE
5	0,94	TAK
…	…	…

Wektory uszeregowane zgodnie z malejącym prawdopodobieństwem przynależności do klasy oznaczonej etykietą *TAK*

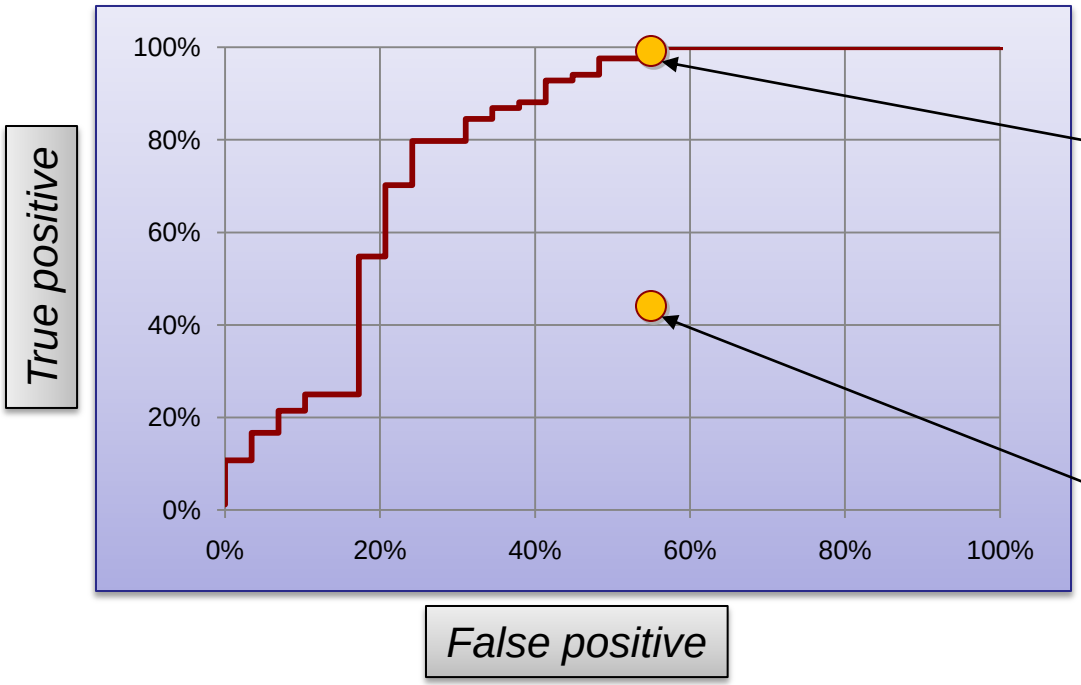


Wystarczy 40% całego zbioru danych, aby znalazło się w nim 80% wszystkich wektorów z klasy „pozytywnej”.

Im większa powierzchnia pod wykresem, tym lepszy klasyfikator.

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Krzywa ROC

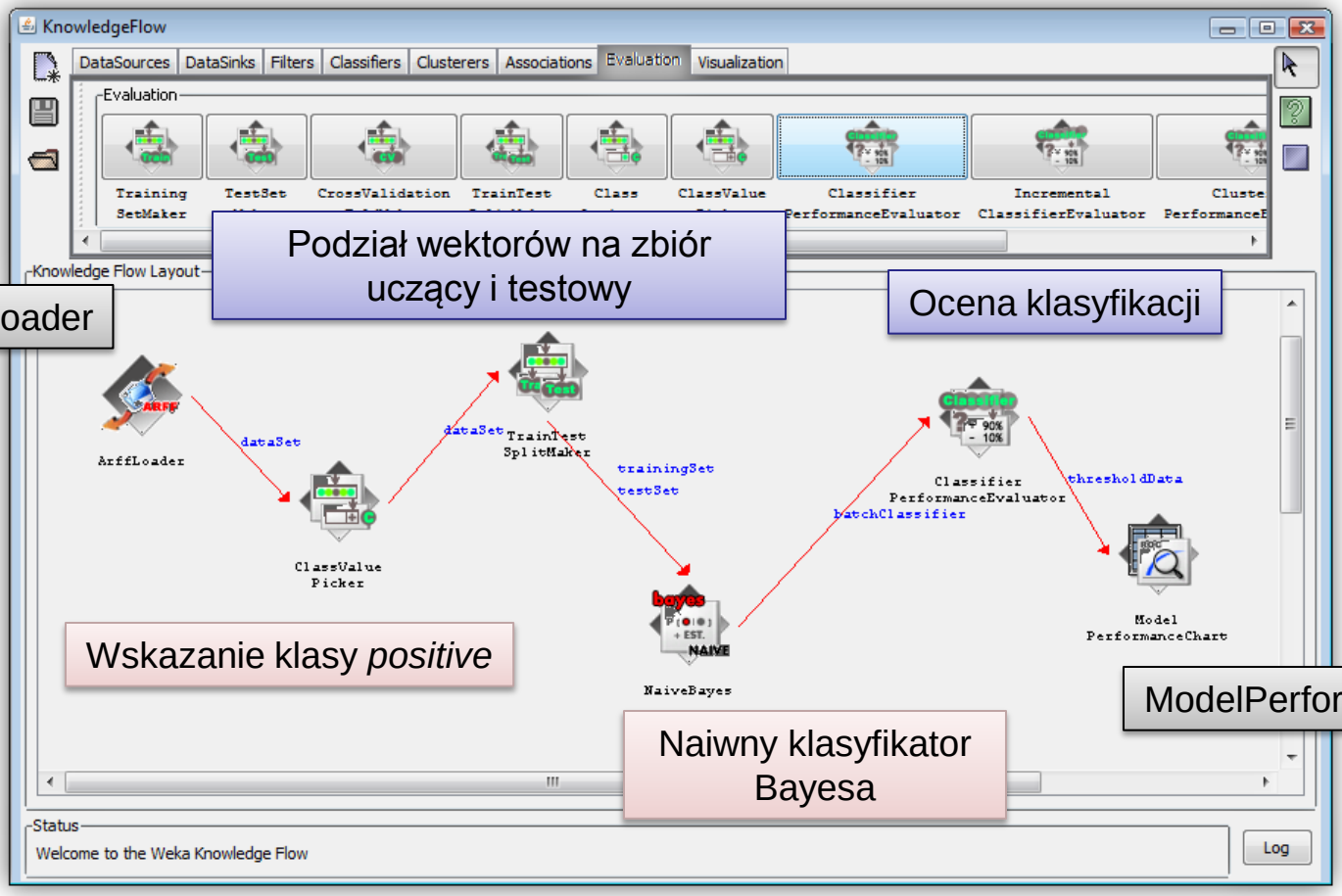


Znalezienie wszystkich wyników typu *TP* odbywa się kosztem popełnienia ok. 55% błędów typu *FP*.

Podobnie jak w przypadku analizy wykresów wzrostu, powierzchnia pod krzywą ROC dla skutecznego klasyfikatora powinna być jak największa.

ROC (ang. *receiver operating characteristic*) — wielkość używana przy detekcji sygnałów transmitowanych przez zaszumiony kanał ; pozwala określić kompromis pomiędzy liczbą właściwych trafień oraz fałszywych alarmów.

Wykonanie krzywej ROC w programie Weka



ArffLoader

Podział wektorów na zbiór uczący i testowy

Ocena klasyfikacji

Wskazanie klasy *positive*

Naiwny klasyfikator Bayesa

ModelPerformanceChart

Prezentacja multimedialna współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego

Ocena wyników wydobywania informacji

W przypadku wydobywania informacji mówi się o dokumentach, a nie o wektorach danych.

Dokumenty zgodne z kryterium wyszukiwania (ang. *relevant documents*) to wektory danych należące do klasy „pozytywnej” (oznaczonej etykietą *TAK*).

$$recall = \frac{\text{liczba wydobytych dokumentów zgodnych z kryterium}}{\text{liczba wszystkich dokumentów zgodnych z kryterium}}$$

$$precision = \frac{\text{liczba wydobytych dokumentów zgodnych z kryterium}}{\text{liczba wszystkich wydobytych dokumentów}}$$

$$F\text{-}measure = \frac{2 \cdot recall \cdot precision}{recall + precision} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$$