

# Miary jakości modelu

Weronika Hryniewska

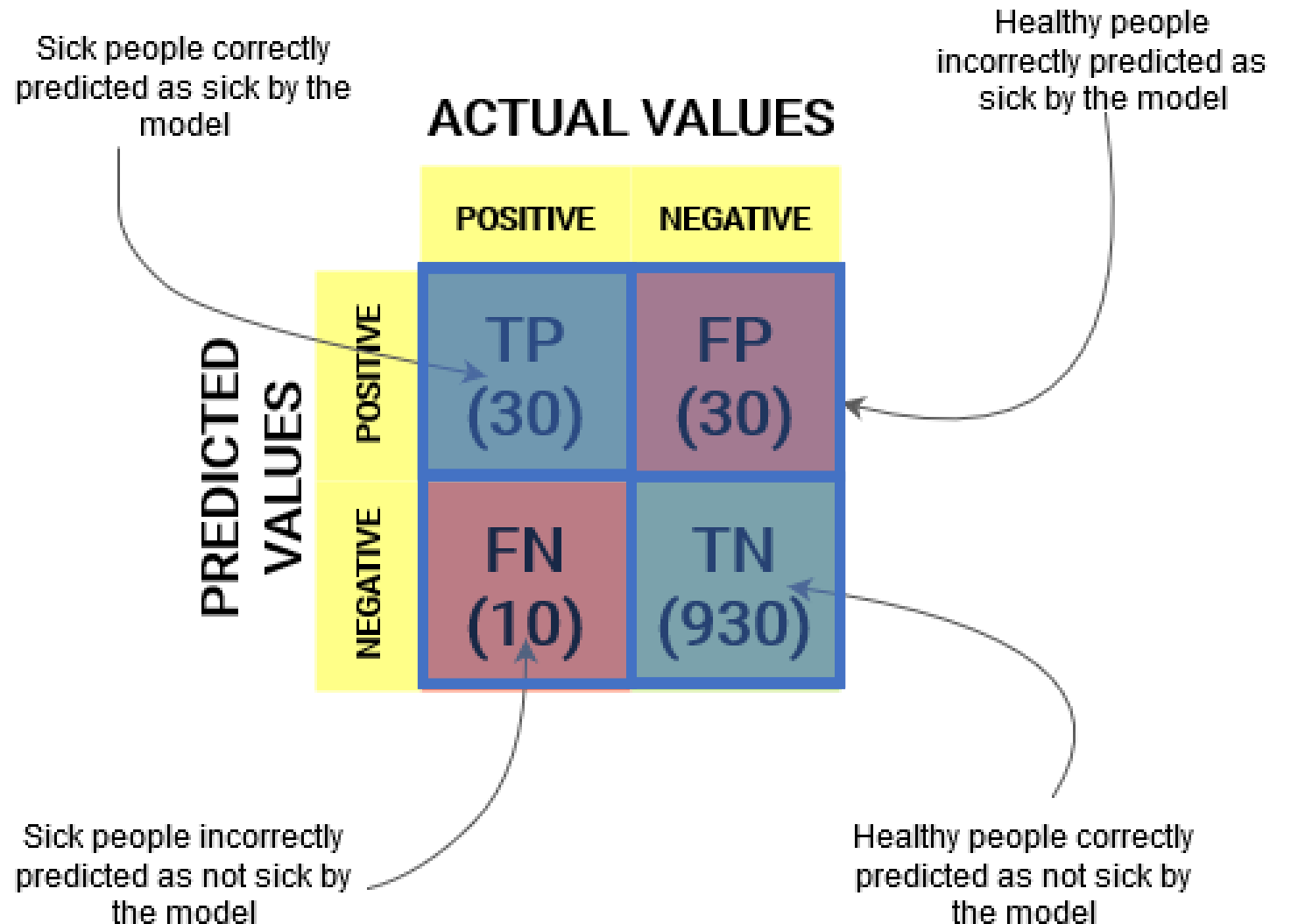
# Confusion matrix

dokładność (ang. accuracy, ACC)

precyzja (ang. precision)

czułość (ang. recall, sensitivity)

swoistość (ang. specificity, SPC)



# Confusion matrix

		True Class	
		Positive	Negative
Predicted Class	Positive	TP prawdziwie dodatnia	FP fałszywie ujemna
	Negative	FN fałszywie dodatnia	TN prawdziwie ujemna



Use Case	High Precision	High Recall	Consequences
Identification of cancer	Desirable	Desirable	<b>Low Precision-</b> unnecessary medical treatment for FP cases <b>Low Recall-</b> undetected cancer patient's treatment get delayed
Identifying a good candidate for hiring	Desirable	Relaxed	<b>Low Precision-</b> Wrong candidate may be chosen <b>Low Recall-</b> Some good candidate may be left behind but the cost of hiring the wrong candidate is more in this case
Predicting Truck Driver Accidents	Relaxed	Desirable	<b>Low Precision-</b> Just extra cost of preventive training to low-risk driver <b>Low Recall-</b> Miss out the accident-prone driver and may end up in major accidents
Predicting stock-outs	Desirable	Relaxed	<b>Low Precision-</b> Getting long list of stocking out SKUs would end in extra cost and measures and may miss out on the actual ones <b>Low Recall-</b> It is ok to have even 60 SKUs out of 200 SKUs so as to at least take preventive measures for these 60 SKUs

Precision-Recall Trade-Off

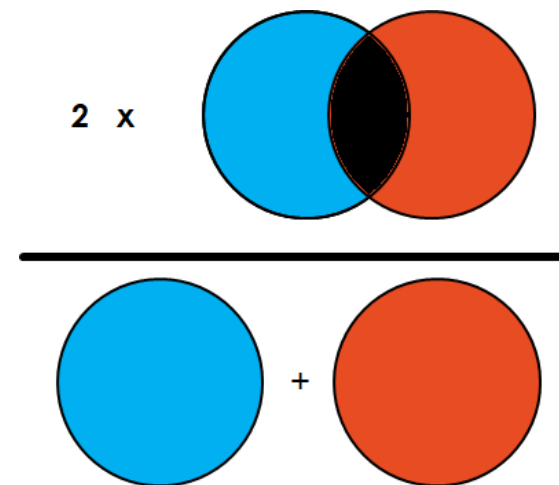
# F1 score

- Średnia harmoniczna precyzji i czułości

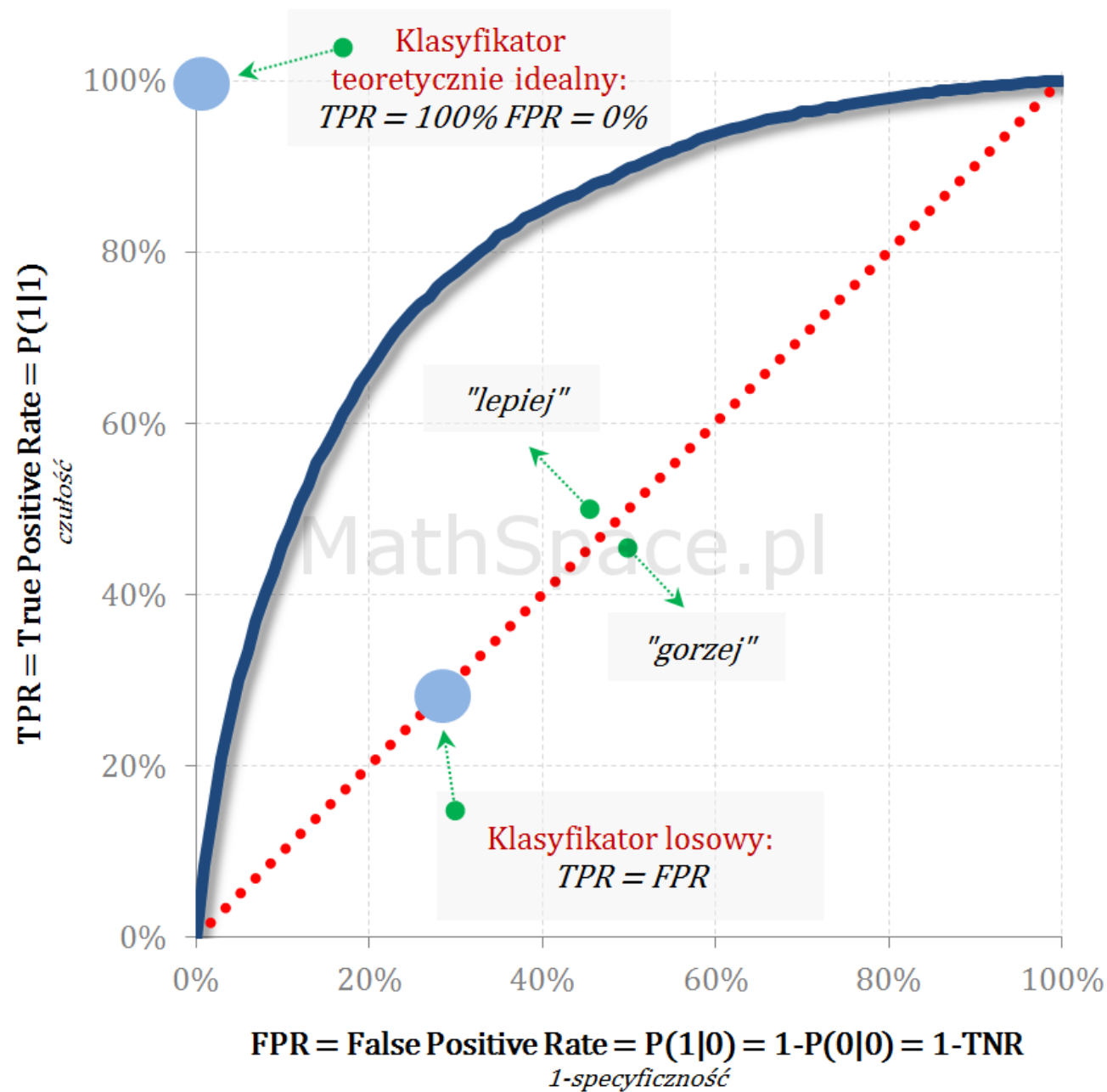
$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

- Kiedy stosujemy?
  - Kiedy FP i FN są tak samo kosztowne - oba wpływają na model prawie tak samo, jak w przykładzie klasyfikacji wykrywania raka
  - Dodanie większej ilości danych nie zmienia efektywnie wyniku
  - TN jest wysoki (jak w przypadku przewidywań nowotworów)

Dice Coefficient



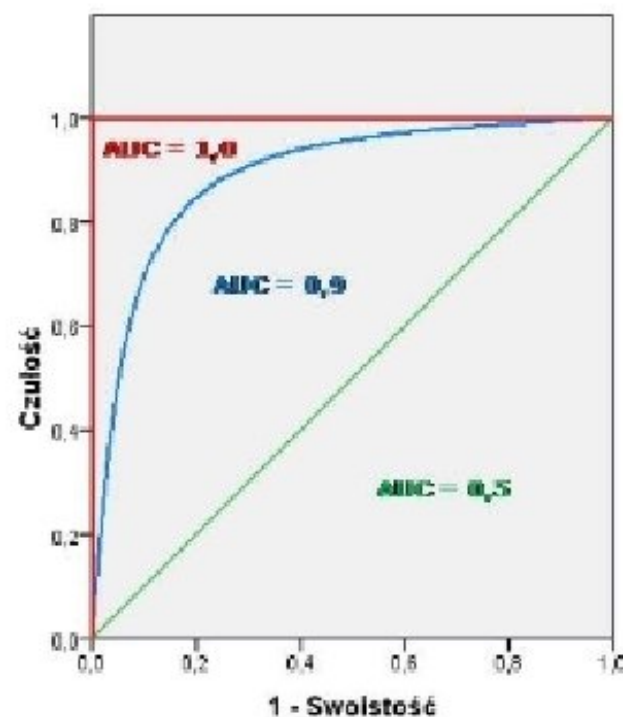
# ROC



# Pole pod wykresem krzywej ROC (AUC)

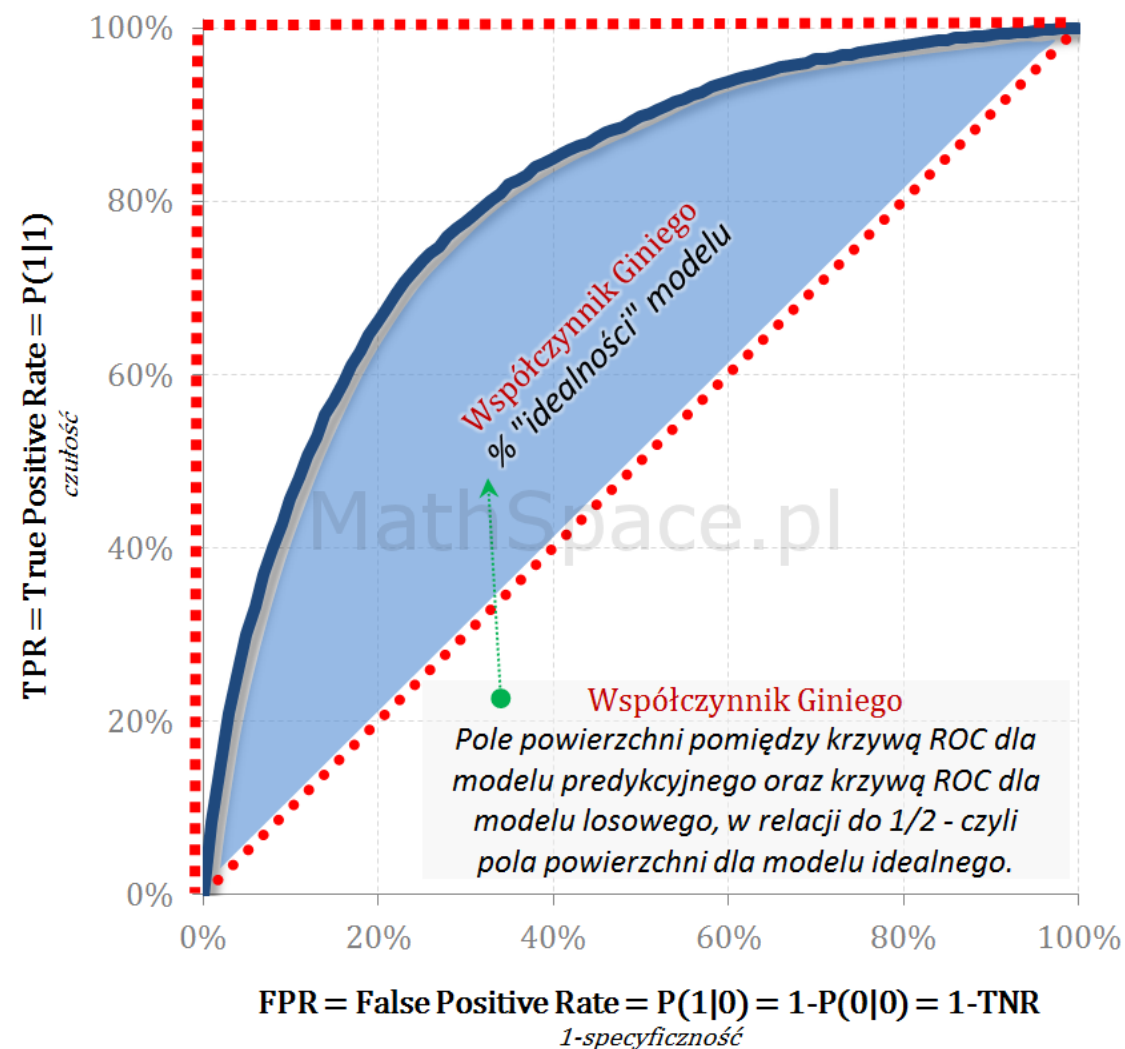
**Pole powierzchni pod wykresem krzywej ROC – AUC** (ang. Area Under Curve), przyjmujące wartości w przedziale od 0 do 1, określa zdolność testu do rozgraniczenia wyników prawidłowych i nieprawidłowych. Im większe jest AUC (im bardziej wklęsła będzie funkcja ROC) tym większa jest moc diagnostyczna testu. Tym samym zdolność rozdzielcza testu jest lepsza.

Krzywa ROC



Większość testów w diagnostyce reprezentuje moc diagnostyczną wyrażającą się wielkościami AUC pomiędzy 0,80 a 0,95.

# Współczynnik Giniego





# Metryki dla klasyfikacji binarnej

- Accuracy
- Precision
- Recall (sensitivity )
- F score
- Specificity
- AUC

# Metryki dla klasyfikacji wieloklasowej

- Average accuracy
- Error rate
- Precision
- Recall
- F score

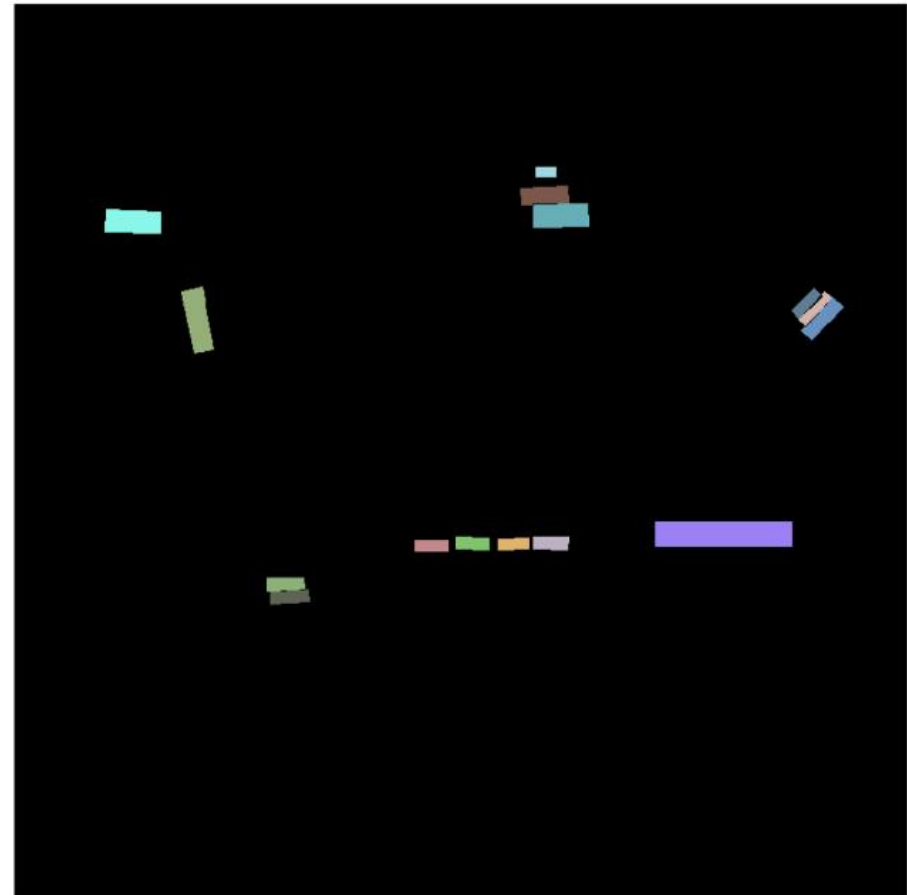
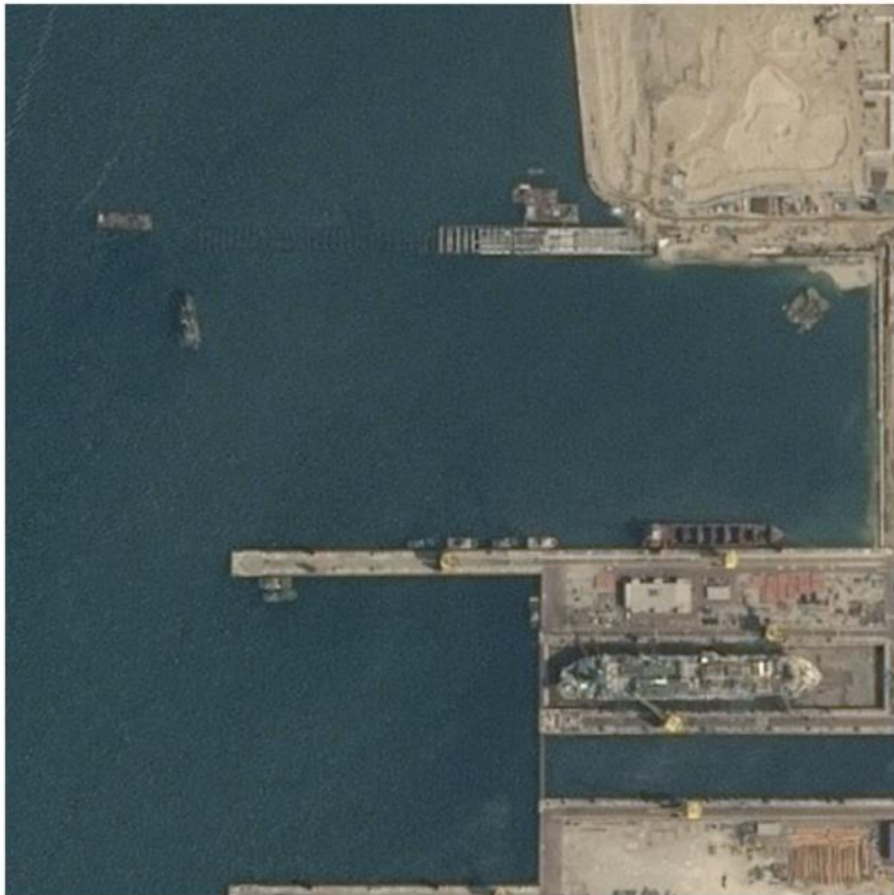
# Metryki dla klasyfikacji wieloetykietowej

- Exact match ratio
- Labelling F score
- Retrieval F score
- Hamming loss
- Precision
- Recall
- F score

# Metryki dla segmentacji

- Sensitivity (Recall)
- Specificity
- F1 score (dice coefficient)
- Pixel accuracy
- IoU, indeks Jaccarda

# Pixel accuracy



# Intersection-Over-Union (IoU, indeks Jaccarda)

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$
