

AKURASI

Akurasi adalah suatu pengukuran untuk mengevaluasi model klasifikasi. Secara informal, akurasi mengukur ketepatan dan kemiripan hasil pada waktu yang sama dengan membandingkannya terhadap nilai absolut.

Beberapa cara untuk pengukuran akurasi diantaranya :

1. Matriks Kumulasi
2. Presisi dan Recall
3. ROC dan AUC

MATRIKS KUMULASI

Definisi True vs False dan Positive vs Negatif

Pada tahap klasifikasi citra, proses evaluasi diperlukan untuk mengetahui bahwa aplikasi yang telah dibuat sudah berjalan secara benar.

Contoh Studi Kasus:

Seorang anak gembala merasa bosan ketika menggembala kambing. Sehingga dia mendapat ide untuk melakukan suatu permainan yaitu dengan berteriak dengan kencang menyebut “Ada SERIGALA !!!” sampai –sampai penduduk desa yang mendengar teriakan itu mendatangi tempat si penggembala meskipun kenyataannya tidak terlihat serigala di tempat itu. Akhirnya penduduk sadar jika mereka dijadikan sebagai bahan candaan dan dibohongi si penggembala.

Hingga suatu malam si penggembala benar-benar melihat serigala mendatangnya. Karena ketakutan, ia berteriak dengan kencang “ada serigala !!!”. Namun penduduk desa terlanjur marah dan tidak mau tertipu lagi yang menyebabkan semua kambingnya habis dimakan serigala lapar.

Dari Contoh cerita di atas bisa kita definisikan sebagai berikut :

- “SERIGALA” sebagai **kelas positive**
- “TIDAK ADA SERIGALA” Sebagai **kelas Negative**

Maka dapat kita simpulkan prediksi keberadaan serigala menggunakan model matrik konfusi 2×2

Yang menggambarkan semua empat kemungkinan hasil output :

True Positive (TP) <ul style="list-style-type: none">• Kenyataannya : Serigala Mengancam• Penggembala berkata : “SERIGALA”• Outcome : Penggembala menjadi Pahlawan	False Positive (FP) <ul style="list-style-type: none">• Kenyataannya : Tidak Ada Serigala• Penggembala berkata : “SERIGALA”• Outcome : Penduduk marah karena dibohongi
False Negative (FN)	True Negative (TN)

<ul style="list-style-type: none"> • Kenyataannya : Serigala Mengancam • Penggembala berkata : “TIDAK ADA SERIGALA” • Outcome : Serigala makan semua kambing 	<ul style="list-style-type: none"> • Kenyataannya : Tidak Ada Serigala • Penggembala berkata : “TIDAK ADA SERIGALA” • Outcome : keadaan aman
---	---

true positive adalah hasil output dimana suatu model klasifikasi secara benar memprediksi kelas positive.

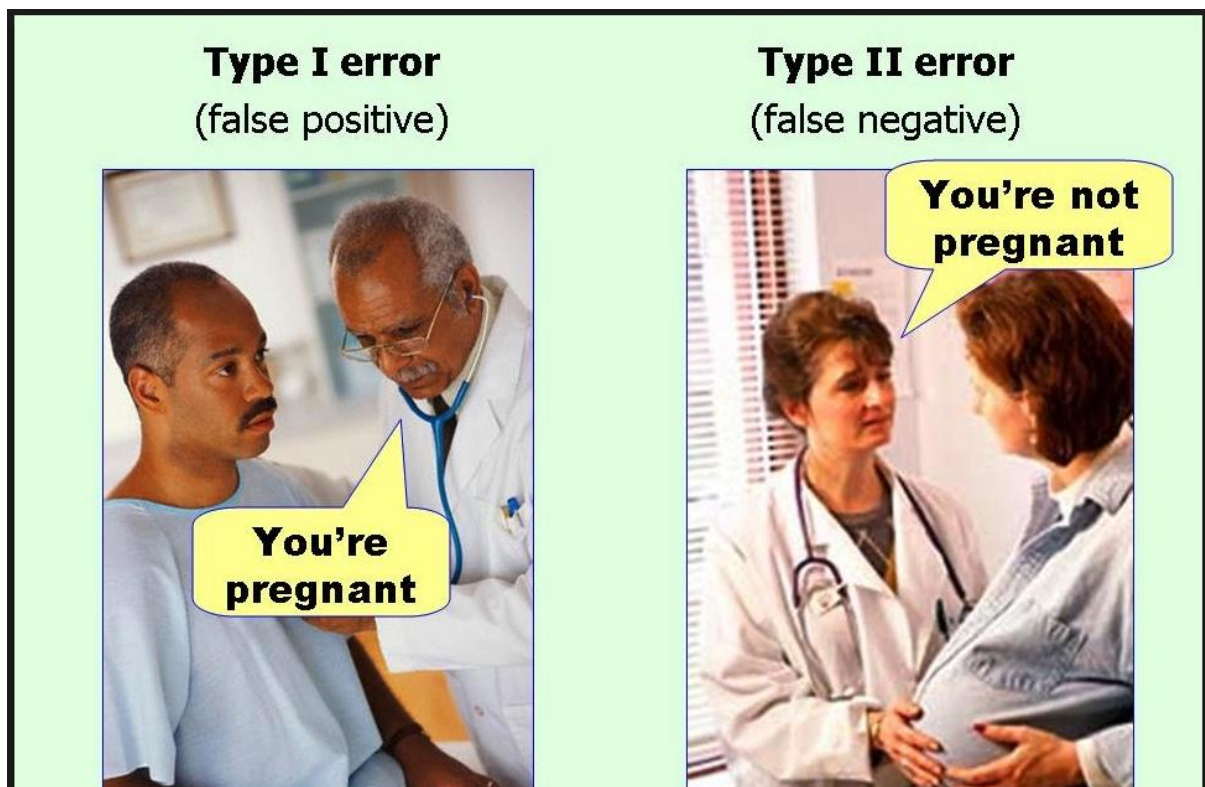
Begitu juga, **true negative** adalah hasil output dimana suatu model klasifikasi secara benar memprediksi kelas *negative*.

false positive adalah hasil output dimana suatu model klasifikasi salah memprediksi kelas positive.

false negative hasil output dimana suatu model klasifikasi salah memprediksi kelas *negative*.

Jika disimbolkan dalam bentuk model biner :

Prediksi dari Model	Apa yg diprediksi
TRUE	POSITIVE
TRUE	NEGATIVE
FALSE	NEGATIVE
FALSE	POSITIVE



CONTOH Studi Kasus : Sistem Deteksi kanker dengan 100 data dimana sistem mendeteksi adanya kanker (kelas Positive) dan tidak adanya kanker (kelas negative).

True Positive (TP) <ul style="list-style-type: none"> • Real : Kanker • Sistem memprediksi : Kanker • Jumlah TP = 1 	False Positive (FP) <ul style="list-style-type: none"> • Real : Tidak ada Kanker • Sistem memprediksi : Kanker • Jumlah FP = 1
False Negative (FN) <ul style="list-style-type: none"> • Real : Kanker • Sistem memprediksi : Tidak ada Kanker • Jumlah FN = 8 	True Negative (TN) <ul style="list-style-type: none"> • Real : Tidak ada Kanker • Sistem memprediksi : Tidak ada Kanker • Jumlah TN = 90

Formulasi dari matrik konfusi adalah sebagai berikut :

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

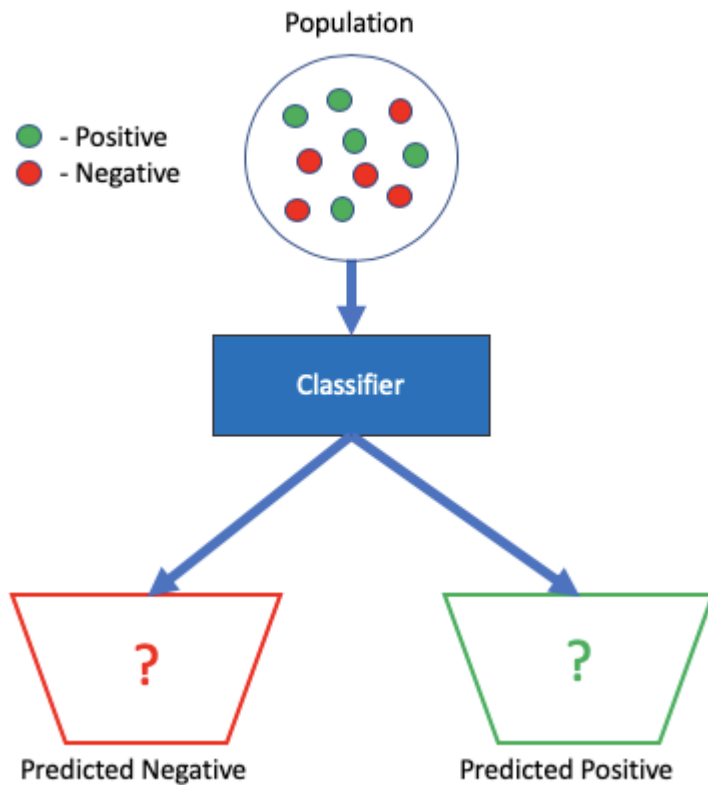
$$Akurasi = \frac{1 + 90}{1 + 90 + 1 + 8} = 0.91$$

Berdasarkan hasil akurasi di atas dapat diambil kesimpulan dari 100 data sample kanker, 91 tidak terdeteksi adanya kanker (90 TN dan 1 FP) dan 9 data terdeteksi sebagai kanker (1 TP dan 8 FN).

Dari 91 data yang bukan kanker, sistem ini telah berhasil mendeteksi 90 data tidak terdapat kanker. Meskipun demikian dari 9 data kanker, sistem ini hanya bisa mendeteksi 1 data yang terindikasi sebagai kanker... Jika sistem ini dimanfaatkan untuk melihat apakah pasien terkena kanker atau tidak maka sistem yang telah dibangun masih kurang bagus karena 8 dari 9 pasien tidak terdiagnosa memiliki penyakit kanker yang menyebabkan keterlambatan pengobatan.

PRESISI dan RECALL

Dalam komunitas data sains,sangat umum melihat system klasifikasi dari segi pengukuran presisi dan recall untuk mengevaluasi model yang dibangun. Sedangkan di dunia medis, umumnya melihat sistem klasifikasi dari segi spesifitas and sensitifitas untuk evaluasi hasil tes medis. Konsep-konsep tersebut hampir mirip namun tetap terdapat perbedaan. Pada saat kedua ilmu bersinggungan dimana tes kesehatan merupakan model Machine Learning, sehingga perbedaan persepsi akan menyebabkan banyak salah faham antara dunia kesehatan dan data.



		Real	
		Positive	Negative
Predicted	Positive	True Positive (tp)	False Positive (fp)
	Negative	False Negative (fn)	True Negative (tn)

$$\text{precision} = \frac{tp}{tp + fp}$$

Precision (Presisi) — Dari keseluruhan sampel yang diprediksi sebagai positive, berapa banyak yang benar-benar positive ?

$$\text{recall} = \frac{tp}{tp + fn}$$

Recall — dari semua sampel positive , berapa banyak yang diprediksi sebagai positive?

$$\text{specificity} = \frac{tn}{tn + fp}$$

Specificity (Spesifitas) — Dari keseluruhan orang yang **tidak sakit**, Berapa banyak yang hasil tesnya negative ?

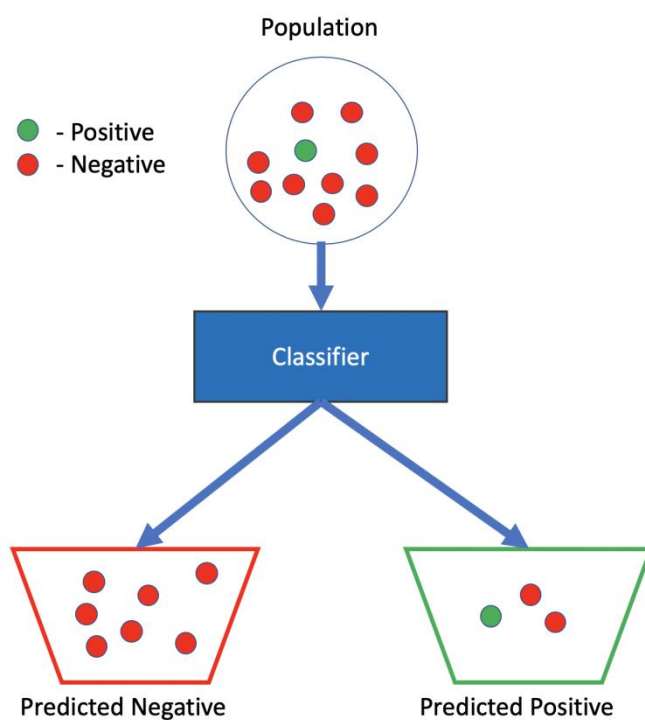
$$\text{sensitivity} = \frac{tp}{tp + fn}$$

Sensitivity (Sensitifitas) — dari keseluruhan orang yang **sakit**, berapa banyak yang hasil tesnya **positive** ?

Jika kita mendefinisikan contoh **positive** sebagai “**orang yang sakit**”, maka definisi recall dan sensitifitas sama. Sedangkan Presisi dan Spesifitas adalah berbeda. Presisi disebut juga *Positive Predictive Value (PPV)*

Implementasi studi kasus permasalahan pengklasifikasian dan model klasifikasi. Masing – masing model klasifikasi mencoba mengelompokkan 10 sampel data ke dalam keranjang *positive* dan *negative* .

Contoh 1 — Low Precision, High Recall, and High Specificity



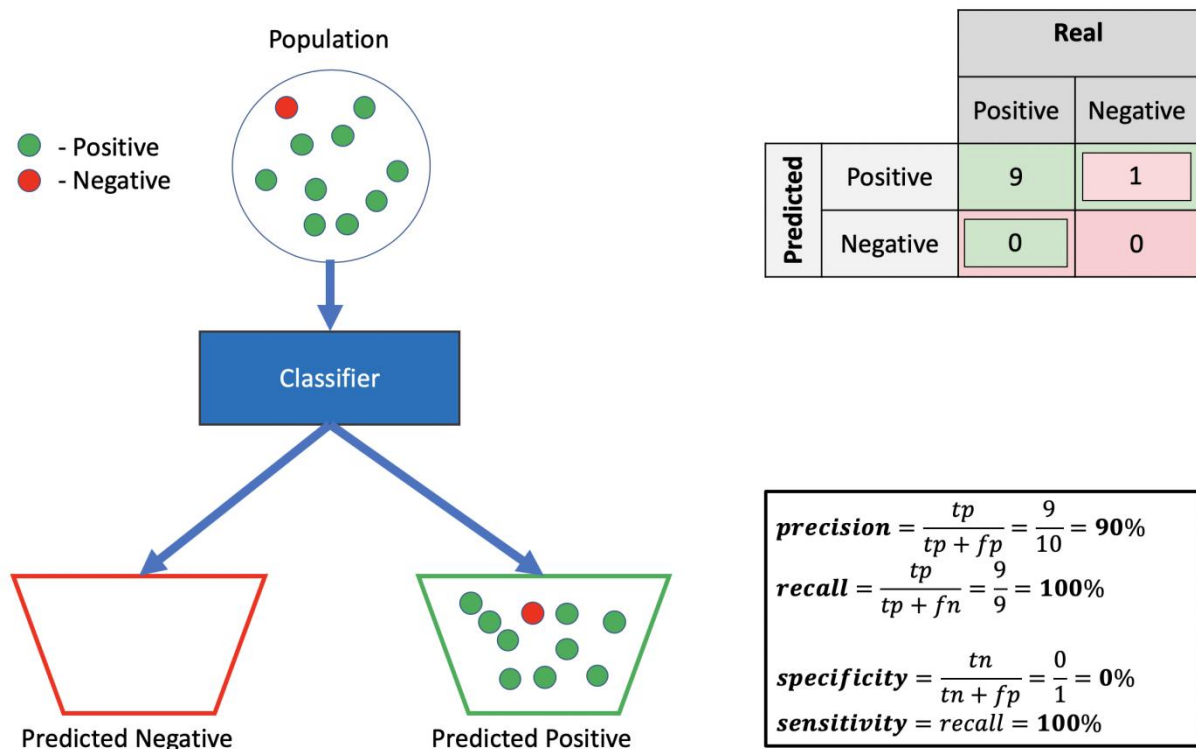
		Real	
		Positive	Negative
Predicted	Positive	1	2
	Negative	0	7

$$\begin{aligned} \text{precision} &= \frac{tp}{tp + fp} = \frac{1}{3} = 33\% \\ \text{recall} &= \frac{tp}{tp + fn} = \frac{1}{1} = 100\% \\ \text{specificity} &= \frac{tn}{tn + fp} = \frac{7}{9} = 78\% \\ \text{sensitivity} &= \text{recall} = 100\% \end{aligned}$$

Jika model klasifikasi **memprediksi negative**, maka hasil dari **pemodelan bisa dipercaya**, sampel data adalah negative. Akan tetapi, harus diingat jika **sampel data adalah negative**, maka **tidak ada jaminan output** dari model diprediksi sebagai negative (specificity=78%).

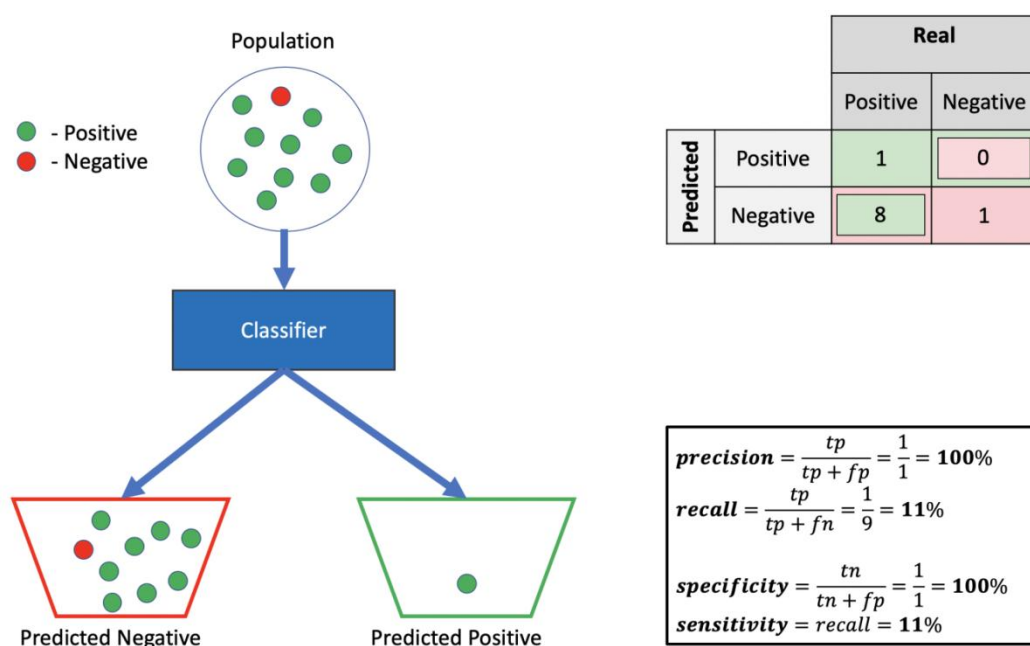
Jika model klasifikasi **memprediksi positive**, maka hasil prediksi tersebut **tidak dapat dipercaya** (precision=33%). Namun, Jika **sampel data adalah positive**, maka hasil output model klasifikasi **dapat dipercaya** (recall=100%).

Contoh 2 — High Precision, High Recall, and Low Specificity



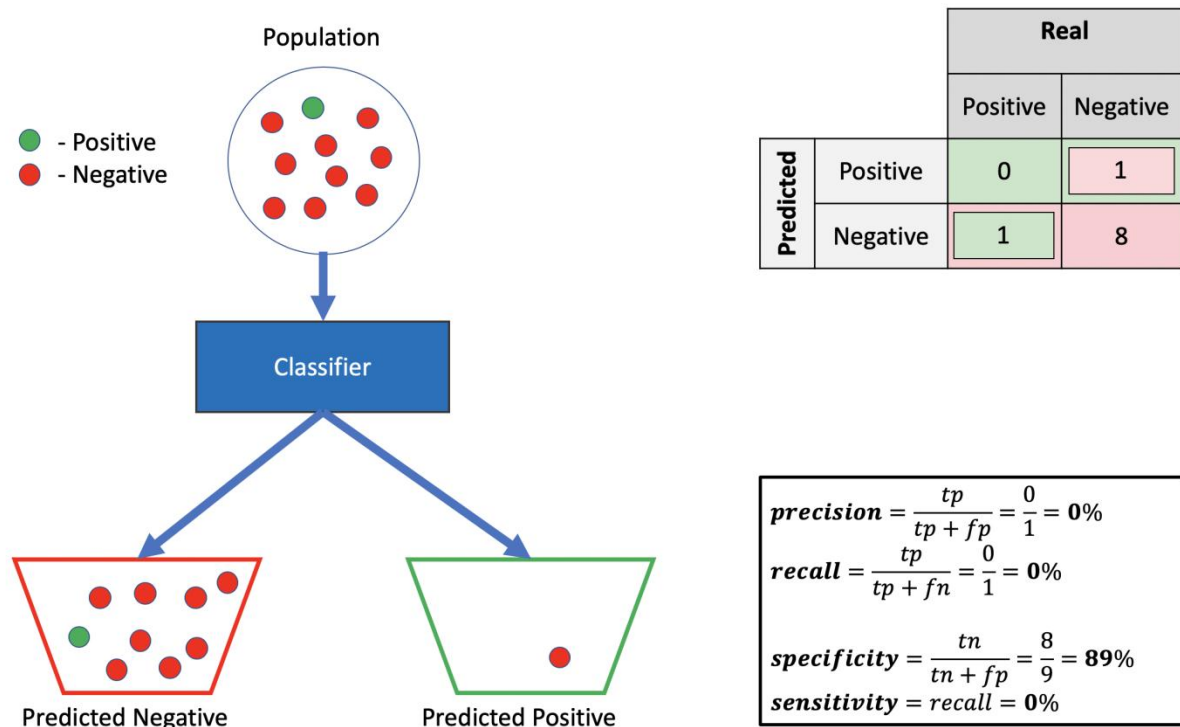
Memprediksi semuanya sebagai positive jelas bukanlah ide yang bagus. Meskipun demikian, populasi yang tidak imbang dimana nilai presisi relatif tinggi yaitu 100 % karena semua sampel positive diprediksi sebagai positive. Di lain sisi, nilai spesifitasnya adalah 0 % karena tidak ada sampel negative yang diprediksi sebagai negative.

Contoh 3 — High Precision, Low Recall, and High Specificity



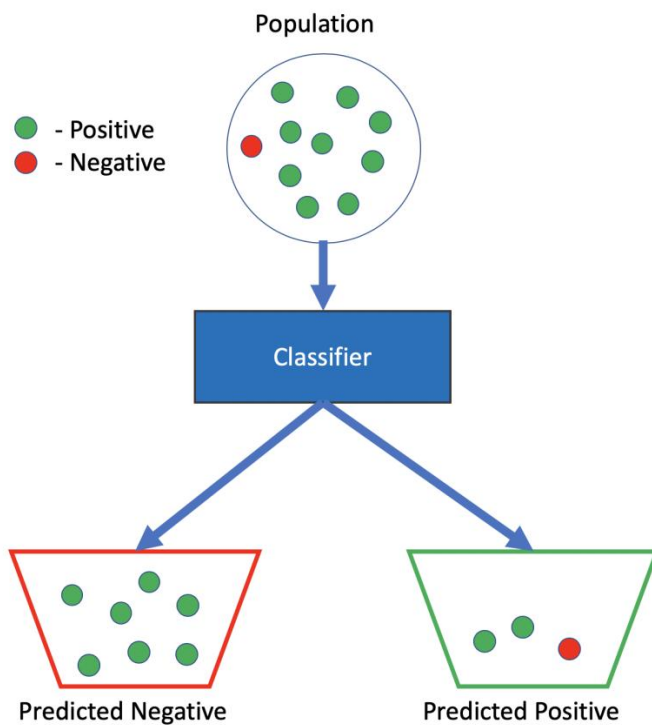
Jika Model Klasifikasi memprediksi sampel data sebagai kelas Positive, maka hasil prediksi dapat dipercaya. Akan tetapi jika hasil prediksi adalah negative, maka hasil prediksi tidak bisa dipercaya sepenuhnya, karena ada kemungkinan hasil prediksi adalah positive.

Contoh 4 — Low Precision, Low Recall, and High Specificity



Model Klasifikasi ini tidak terlalu bagus — semua prediksi mengarah ke kelas negative, dan saat hasil prediksinya adalah positive — maka hasil tersebut adalah salah. Sebenarnya penggunaan yang berkebalikan dengan yang diprediksi model akan menghasilkan nilai yang lebih baik.

Contoh 5 — High Precision, Low Recall, and Low Specificity



		Real	
		Positive	Negative
Predicted	Positive	2	1
	Negative	7	0

$$\text{precision} = \frac{tp}{tp + fp} = \frac{2}{3} = 66\%$$

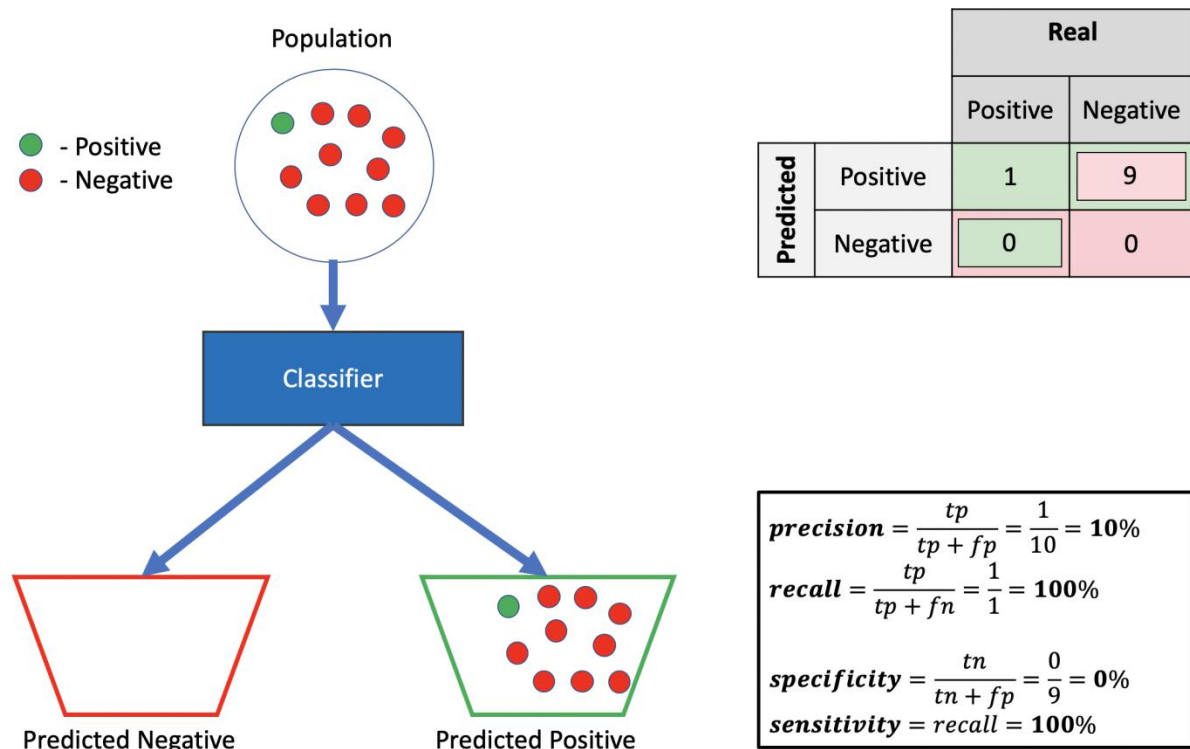
$$\text{recall} = \frac{tp}{tp + fn} = \frac{2}{9} = 22\%$$

$$\text{specificity} = \frac{tn}{tn + fp} = \frac{0}{1} = 0\%$$

$$\text{sensitivity} = \text{recall} = 22\%$$

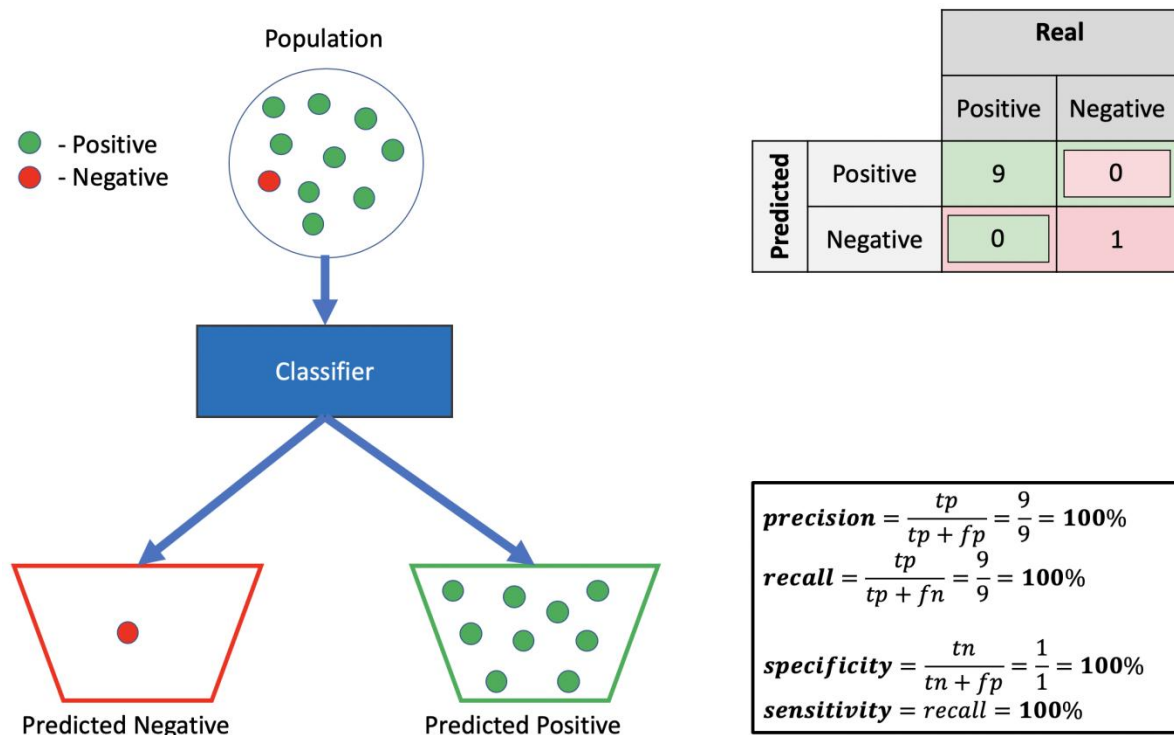
Sama halnya dengan contoh 4. Melakukan klasifikasi dengan hal yang berkebalikan dari hasil prediksi model klasifikasi akan menghasilkan kinerja yang lebih baik

Contoh 6 — Low Precision, High Recall, and Low Specificity



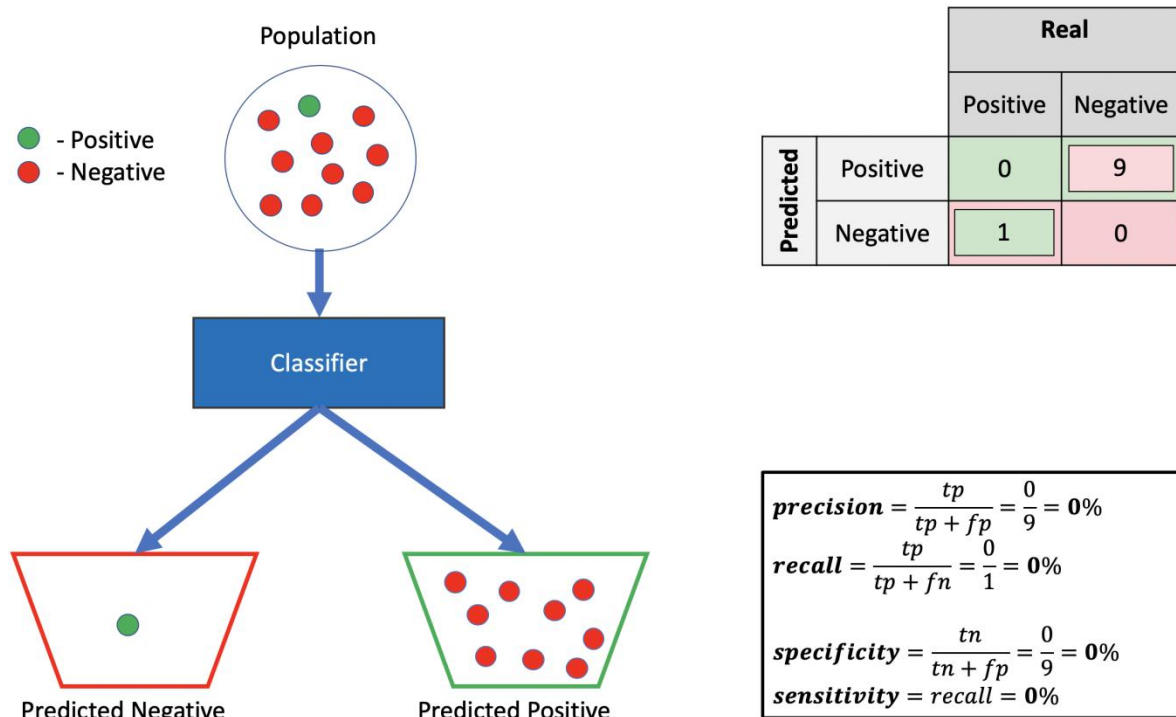
Model klasifikasi semacam ini kemungkinan tidak berguna dimana model ini memprediksi semuanya sebagai positive dan tentu saja model ini secara sempurna berhasil mendeteksi sampel data positive, akan tetapi kita tidak mendapatkan informasi apapun yang berguna untuk penelitian kita.

Contoh 7 — High Precision, High Recall, and High Specificity



Model klasifikasi yang sangat sempurna — Model ini mendeteksi semua sampel data positive sebagai positive dan sampel data negative sebagai negative. Semua pengukuran bernilai sempurna yaitu 100 %.

Contoh 8 — Low Precision, Low Recall, and Low Specificity



Model klasifikasi yang buruk. Semua sampel data positive diprediksi sebagai negative. Begitu juga semua data negative diprediksi sebagai positive. Semua pengukuran bernilai 0 %. Namun secara sederhana bisa gunakan untuk memprediksi kebalikannya sehingga bisa menjadi model klasifikasi yang sempurna.

Kesimpulannya, semua bentuk pengukuran (Presisi, Recall dan Spesifitas memberikan informasi yang penting tentang seberapa baiknya kinerja model kasifikasi. Meskipun dengan hanya Recall dan Presisi sudah cukup, akan tetapi pengukuran specifitas model diperlukan terutama pada dunia medis

ROC plots 2 parameters:

1. True Positive Rate (TPR)

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

2. False Positive Rate (FPR)

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

Saat membuat kurva ROC, maka plot 'True Positives' di Y-axis & 'False Positives' di X-axis.

Studi Kasus :

Kita asumsikan memiliki 200 data dengan 100 data diklasifikasikan sebagai kelas "0" dan sisanya masuk ke kelas "1"

Setelah dilakukan tahap prediksi pada model klasifikasi ini , maka didapat hasil sebagai berikut :

Jumlah data yang masuk kelas "0" = 30 data

Jumlah data yang masuk kelas "1" = 70 data

Sehingga didapatkan nilai *true positive rate* (TPR) 0.70 dalam artian $\frac{70}{100}$ dalam prosentase = 70%

Sedangkan 100 data lainnya memiliki hasil tes klasifikasi sebagai berikut :

Jumlah data yang masuk kelas "1" = 30 data

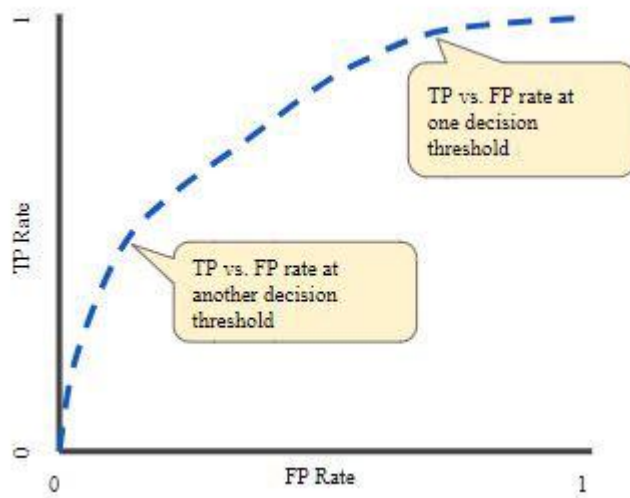
Jumlah data yang masuk kelas "0" = 70 data

Sehingga didapatkan nilai *false positive rate* (FPR) 0.70 dalam artian $\frac{70}{100}$ dalam prosentase = 70%

Yang menjadi pertanyaan adalah apa yang akan terjadi jika TPR = FPR

Sesuai hasil prediksi, TPR untuk kelas "1" 70, begitu juga FPR untuk kelas "0" adalah 70 sehingga totalnya menjadi 140 diman 50% adalah benar (*true positive*) dan 50% adalah tidak benar (*false positive*).

Kurva untuk FPR vs TPR diplot dengan threshold berbeda menggunakan kurva ROC. Jika nilai threshold direndahkan, maka model akan lebih banyak mengklasifikasikan sebagai positive

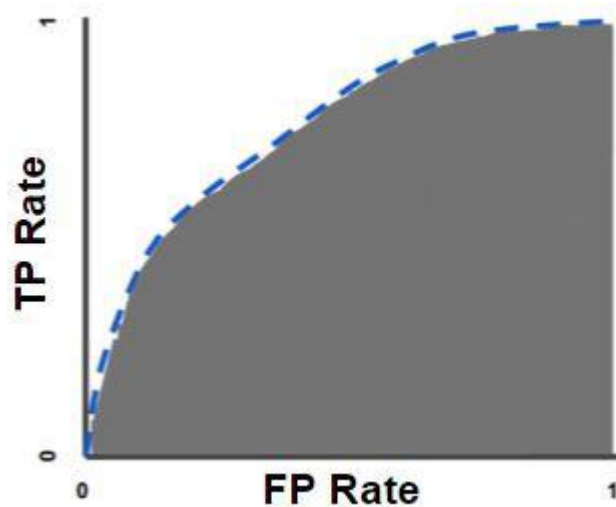


contoh Thresholds

Selanjutnya jika kita akan menghitung titik pada kurva ROC, kita bisa mengevaluasinya dengan Regresi logistik menggunakan threshold yang berbeda untuk beberapa kali. Akan tetapi hal ini tidak efisien. Sehingga untuk pengukuran kualitas klasifikasi bisa menggunakan Algoritma AUC.

AUC: Area Under The Curve/ ROC Curve

Model ini mengukur seluruh area 2D dari sumbu x dan sumbu y dibawah kurva ROC dari (0,0) sampai (1,1).



Kurva AUC

AUC membantu menghasilkan pengukuran agregate untuk mengetahui kinerja dari keseluruhan threshold klasifikasi. AUC range bernilai 0 - 1.

Model yang memiliki nilai prediksi 100 % salah maka nilai AUC adalah 0.0 dimana dengan model yang mirip yang mprediksi 100 % benar memiliki AUC sebesar 1.0

Sehingga dapat ditarik kesimpulan :

Model klasifikasi yang sangat lemah memiliki AUC kisaran 0.5 & model klasifikasi yang baik memiliki nilai AUC pada kisaran 1