

F-MEASURE

WHAT IS THIS COURSE ABOUT?

- Recall
- Precision
- Accuracy
- F-Measure
- Contoh Penyelesaian

PENGANTAR

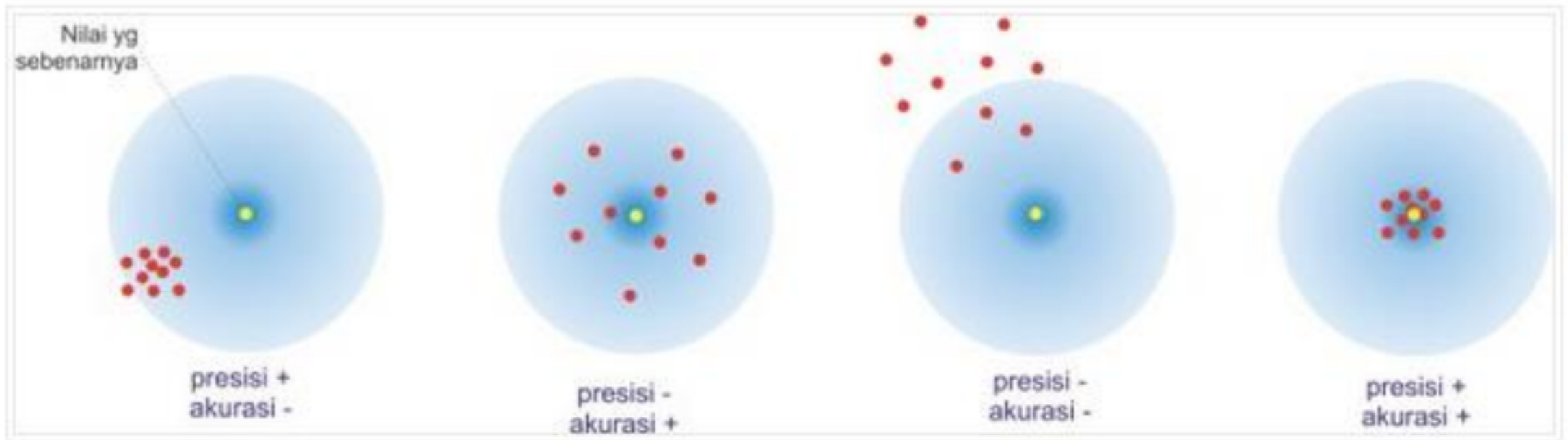
Dalam “dunia” pengenalan pola (*pattern recognition*) dan temu kembali informasi (*information retrieval*), *precision* dan *recall* adalah dua perhitungan yang banyak digunakan untuk mengukur kinerja dari sistem / metode yang digunakan.

Precision adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem.

Recall adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi.

Sedangkan di dunia statistika dikenal juga istilah *accuracy*. **Accuracy** didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual.

PENGANTAR (2)



Accuracy vs Precision [4]

CONTOH

Sebuah formulasi untuk pengelompokan spam email secara otomatis akan diuji kinerjanya, menggunakan **100** konten yang terbukti spam dan **900** konten bukan spam. Setelah proses formulasi dijalankan ternyata terdeteksi ada **110** konten spam. Setelah dilakukan pengecekan dengan konten asli ternyata dari **110** konten yang dideteksi spam ada **20** konten yang ternyata bukan spam.

Dari kasus tersebut maka kita dapat simpulkan bahwa formulasi tersebut memiliki **precision** sebesar 82%, **recall** sebesar 90% dan **accuracy** sebesar 97% .

CONTOH (2)

Precision : $\frac{\text{Jumlah konten spam yang dinyatakan benar}}{\text{Jumlah konten yang dideteksi spam}}$

$$\text{Precision} : \frac{90}{110} = 0,82 = 82 \%$$

Recall : $\frac{\text{Jumlah konten spam yang dinyatakan benar}}{\text{Jumlah konten spam sebenarnya}}$

$$\text{Recall} : \frac{90}{100} = 0,9 = 90 \%$$

Accuracy : $\frac{\text{Jumlah konten (spam+non spam) yang dideteksi dan dinyatakan benar}}{\text{Jumlah total konten}}$

$$\text{Accuracy} : \frac{90+880}{1000} = 0,97 = 97 \%$$

RUMUS

		Nilai sebenarnya	
		TRUE	FALSE
Nilai prediksi	TRUE	TP (True Positive) <i>Corect result</i>	FP (False Positive) <i>Unexpected result</i>
	FALSE	FN (False Negative) <i>Missing result</i>	TN (True Negative) <i>Corect absence of result</i>

$$precision = \frac{TP}{TP + FP}$$

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

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

KASUS SPAM

		Nilai sebenarnya	
		TRUE	FALSE
Nilai predisksi	TRUE	90	20
	FALSE	10	880

$$\text{precision} = \frac{90}{90 + 20} = \frac{90}{110} = 0.82 = 82\%$$

$$\text{recall} = \frac{90}{90 + 10} = \frac{90}{100} = 0.9 = 90\%$$

$$\text{accuracy} = \frac{90 + 880}{90 + 880 + 20 + 10} = \frac{970}{1000} = 0.97 = 97\%$$

KASUS SPAM (2)

Menggunakan *precision* atau *accuracy* saja dalam sebuah mengukur kinerja dari sebuah sistem / metode bisa menimbulkan bias yang sangat fatal. Sebagai contoh, dari kasus deteksi spam diatas misalkan formulasi menghasilkan hanya ada 1 dokumen spam dan setelah dicek itu benar dokumen spam, maka dihasilkan recall precision sebagai berikut:

Dari hasil perhitungan kita dapatkan precision sebesar 100% dan accuracy sebesar 90.1%. Sekilas tampak baik, namun perhatikan nilai recall yang hanya sebesar 1%. Hal ini menunjukkan bahwa sistem hanya dapat memisahkan konten spam dalam jumlah yang sedikit sekali dan masih banyak konten spam yang tidak diakui.

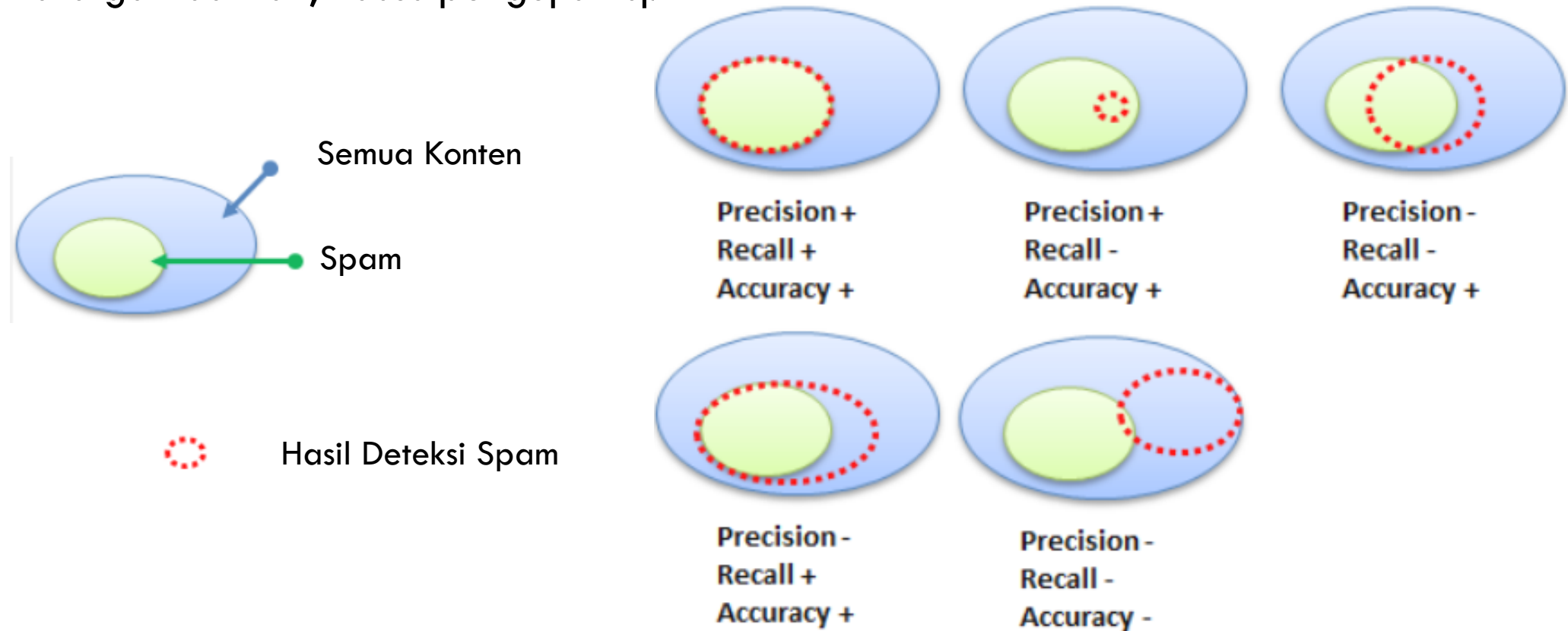
		Nilai sebenarnya	
		TRUE	FALSE
Nilai prediksi	TRUE	1	0
	FALSE	99	900

$$precision = \frac{1}{1 + 0} = \frac{1}{1} = 1 = 100\%$$

$$recall = \frac{1}{1 + 99} = \frac{1}{100} = 0.01 = 1\%$$

KASUS (3)

Bila digambarkan, kasus pengujian spam tadi dapat digambarkan seperti ini:



KESIMPULAN

Dalam mengukur kinerja dari sebuah sistem / metode dalam pengenalan pola atau temu kembali informasi disarankan menggunakan minimal dua parameter yaitu *precision* dan *recall* untuk mendeteksi bias seperti pada kasus diatas.

F-MEASURE

Rumus F-measure (Penggabungan recall dan precision)

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