

Per-Document Evaluation

	c_1	c_2	c_3	\dots	c_k
d_1	y(+)	y(-)	n(+)		n(+)
d_2	y(-)	n(+)	y(+)		n(+)
d_3	n(+)	n(+)	y(+)		n(+)

How good are the decisions on d_i ?

When the system says "yes,"
how many are correct?

↓

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

↗

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

	System ("y")	System ("n")
Human (+)	True Positives TP	False Negatives FN
Human (-)	False Positives FP	True Negatives TN

Does the doc have all the categories
it should have?