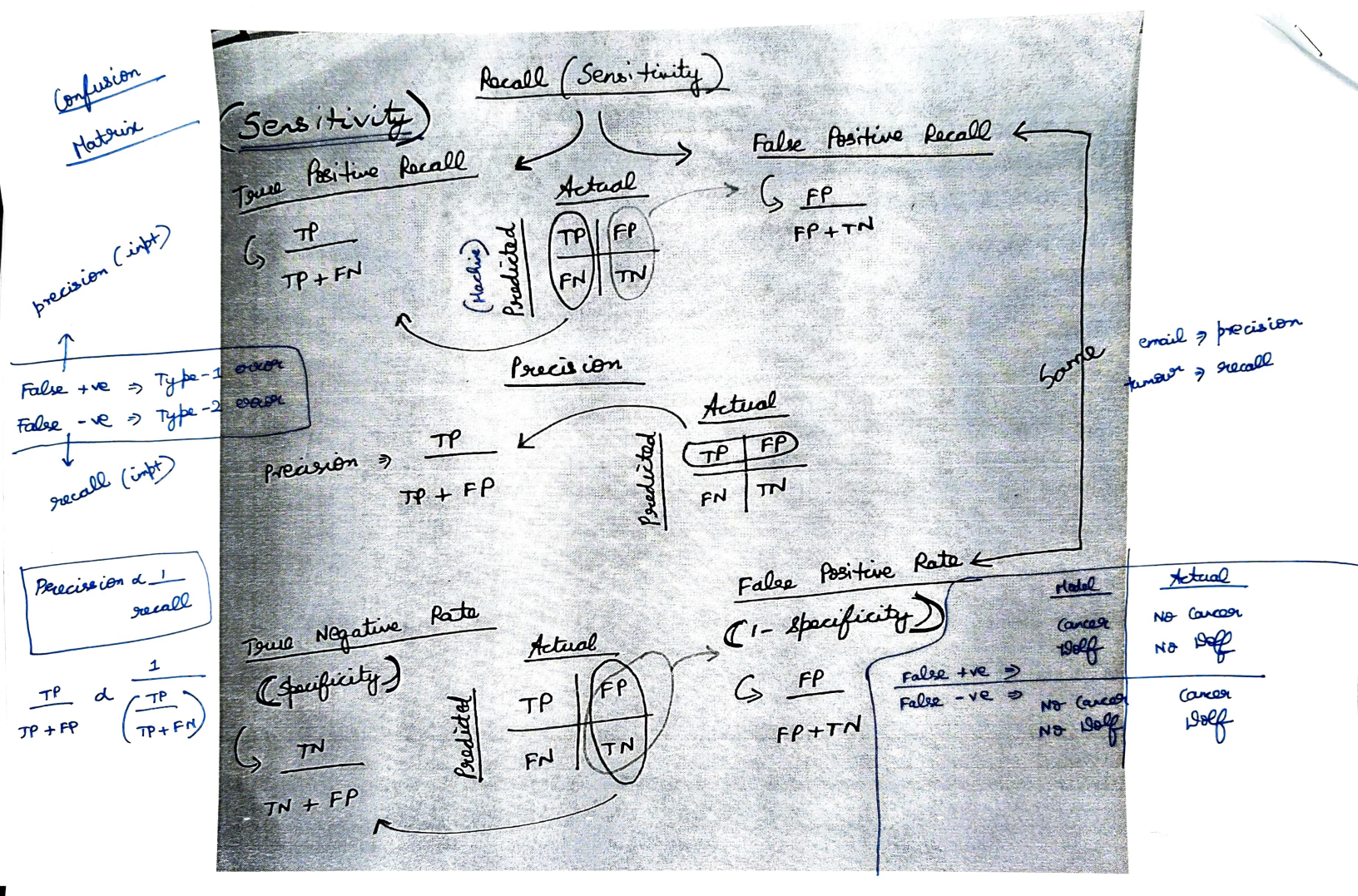

Confusion Matrix
Recall (Sensitivity)
(Sensitivity)
True Positive Recall
TP / TP + FN
False Positive Recall
FP / FP + TN
Actual
Predicted (Machine)
TP
FP
FN
TN
Precision
Precision ⇒ TP / TP + FP
Some
email ⇒ precision
tumour ⇒ recall
precision (inpt)
False +ve ⇒ Type-1 error
False -ve ⇒ Type-2 error
recall (inpt)
Precision α 1 / recall
TP / JP + FP α 1 / (TP / TP + FN)
True Negative Rate
(Specificity)
TN / TN + FP
False Positive Rate
(1 - Specificity)
FP / FP + TN
Model
Actual
Cancer
Daff
NO Cancer
NO Daff
False +ve ⇒
False -ve ⇒

F1 measure ⇒ $\frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}}$

[It uses Harmonic Mean in place of Arithmetic Mean by punishing the extreme values more]

↓

More important than accuracy with mean class distribution

Accuracy

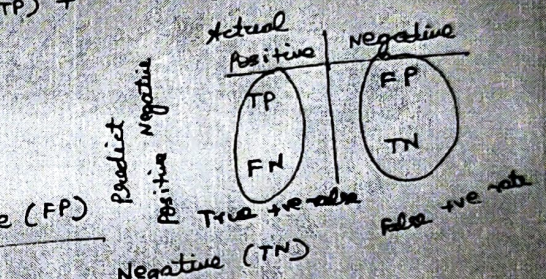
$$\text{Accuracy} \Rightarrow \frac{\text{True Positive (TP)} + \text{True Negative (TN)}}{TP + TN + FP + FN}$$

Recall / True Positive Recall :

$$\text{Recall} \Rightarrow \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Negative (FN)}}$$

Recall / False Positive Recall :

$$\text{Recall} \Rightarrow \frac{\text{False Positive (FP)}}{\text{False Positive (FP)} + \text{True Negative (TN)}}$$



Balanced vs Imbalanced data-sets

	1 (true)	0 (false)
Balanced data-sets	1000	900
Imbalanced data-sets	1000	100

↳ fraud detection
fraud transaction

* We will use Up Sampling and create points.
* Every disease related data-sets are imbalanced since everybody is not having disease.