Performance Matrices

Dataset :

x_1	x_2	y	$\hat{y}$
—	—	0	1
—	—	1	1
—	—	0	0
—	—	1	1
—	—	1	1
—	—	0	1
—	—	1	0

* **Confusion Matrix** : ⟶ Matrix

		Actual 1	Actual 0
Predicted	1	TP	FP
	0	FN	TN

		Actual 1	Actual 0
Predicted	1	3	2
	0	1	1

* **Accuracy** :

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

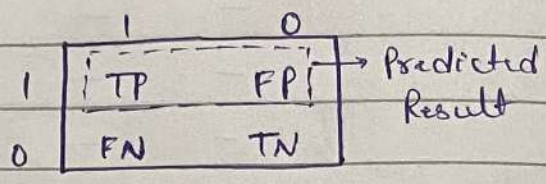
$$\Rightarrow A = \frac{3+1}{3+1+2+1} = \frac{4}{7} = 0.57$$

Note : Dataset → {0 : 900, 1 : 100} → Imbalanced Dataset
For imbalanced dataset, accuracy can be misleading, so we should not depend on accuracy. Either we should use Precision & Recall.

* Precision :

It measures that out of all the predicted results, how many of them are correctly predicted.

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

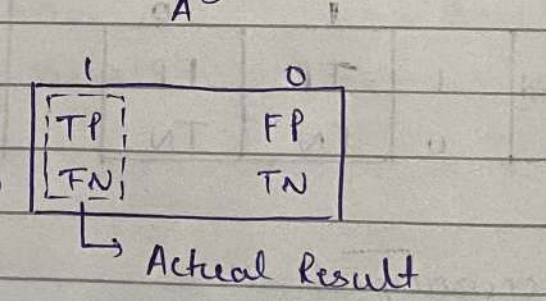


Note: In precision, we reduce false positive (FP)

* Recall :

It measures that out of all the actual results, how many of them are correctly predicted.

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



Note: In recall, we reduce false negative (FN).

Q when to use precision and when to use recall?

→ Based on scenarios, we should use precision and recall.

• Spam classification :

① Mail → Spam
Model → Not Spam
} → Still Good

② Mail → Spam } → Good
Model → Spam

		A: S	A: NS
P	S	TP	FP
	NS	FN	FP

③ Mail → Not spam } → Blunder
Model → Spam

In this case, we will reduce false positive (FP), therefore we will use precision.

- Person has cancer or not :

① Person → cancer } → Blunder
Model → Not cancer

② Person → cancer } → Good
Model → cancer

		A: C	A: NC
P	C	TP	FP
	NC	FN	TN

③ Person → Not cancer } → Still good
Model → cancer

In this case, we will reduce false negative (FN), therefore we will use recall.

Q what to do when false positive (FP) and false negative (FN), both are important ?

→ Then, we will use another metric which is F-Beta score.

* F-Beta Score :

$$\text{F-Beta Score} = \frac{(1+\beta)^2 (P \times R)}{\beta^2 \times [P+R]}$$

- P → Precision
- R → Recall

→ Case 1 : FP and FN, both are important

$\beta = 1 \Rightarrow \frac{(1+1)(P \times R)}{(P+R)}$ → Harmonic Mean

→ Case 2 : FP is more important than FN

$\beta = 0.5 \Rightarrow \frac{(1+0.25)(P \times R)}{0.25[P+R]}$

→ Case 3 : FN is more important than FP

$\beta = 2 \Rightarrow \frac{(1+4)(P \times R)}{4[P+R]}$