Classification Matrics (PRECISION)

There are 2 models to detect email is spam or not. Both model have some accuracy. Which model should be selected

Model A →

		Prediction	
		Spam	Not spam
Actual	Spam	100	170 (FN)
	Not spam	20 (FP)	700

Model B →

		Prediction	
		Spam	Not spam
Actual	Spam	100	190 (FN)
	Not spam	10 (FP)	700

→ False positive (Type I error)

- When model predicted email to be spam, but actually it was not.

→ False Negative (Type II error)

- When model predicted email not spam, but it was actually a spam

- Identify which error is dangerous

- Example : • company sent you mail regarding, that you are placed.

  • But model sent it to spam. but it was not a spam.

♡

∴ False positives are more dangerous in this situation.

→ Hence, you will select a model which have less False positive error ie model B will be selected.

→ This is called precision.

→ Precision: What proportion of predicted positives are actually positives?

$$= \frac{TP}{(TP+FP)} \quad \begin{array}{l} \rightarrow \text{Actual positives} \\ \rightarrow \text{Predicted positives} \end{array}$$

→ Model's A precision $= \dfrac{100}{100+30} = \dfrac{100}{130} = 0.7$

→ Model's B precision $= \dfrac{100}{100+10} = \dfrac{100}{110} = 90$

• Since Model's B precision is more, model B should be selected in this situation.

• High precision indicates, less false positive errors.

RECALL

(1) You are given 2 models that detect cancer. Both model have same accuracy. You are assigned to choose the model.

1) Model A →

		Prediction	
		Detected Cancer	Not detected
Actual →	has cancer	1000	200 (F.N)
	No cancer	800 (F.P)	8.000

2) Model B →

		Predicted	
		Detected cancer	Not detected
Actual	has cancer	1000	500 (F.N)
	No cancer	500 (F.P)	8000

→ False positive
- model predicted cancer, when person doesn't have cancer.

→ False negative
- model predicted no cancer when person actually has a cancer.

→ Identify which error is more dangerous.

- In this situation, False negatives are more dangerous ie model doesn't detect cancer, when it actually has cancer.

- Hence, we will select that model which has less False negative error

- Since Model A has less False Negative, it will be selected.

- This is called Recall.

- Recall :- What proportions of positives are are correctly classified.

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

→ Model A's recall $= \dfrac{1000}{1000 + 200} = 0.83$

→ Model B's recall $= \dfrac{1000}{1000 + 500} = 0.66$

- Since, model's A recall is more, Model's A will be select.

- Higher Recall indicates, less False negatives