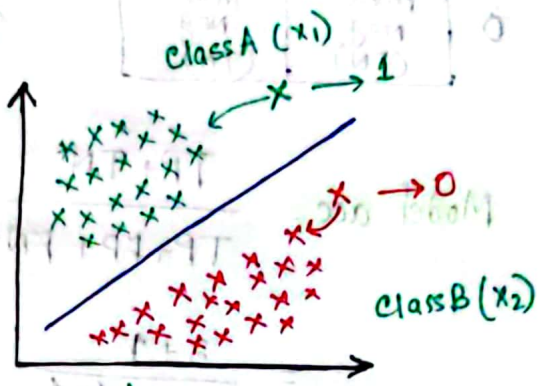
Processes after model Training: (Finding model accuracy)

1. Confusion Matrix
2. ~~Accuracy~~
3. Precision (Acc tecnique 1)
4. Recall (Acc tecnique 2)
5. F$_\beta$ - Beta Score (Acc technique 3)



Dataset

x_1	x_2	y (Actual Value)	$\hat{y}$ (Predicted Value)
–	–	0	1
–	–	1	1
–	–	0	0
–	–	1	1
–	–	1	1
–	–	0	1
–	–	1	0

y $\hat{y}$
0 - 0 → Correct prediction
1 - 1 → Correct "
1 - 0 → Not correct "
0 - 1 → Not Correct "

Confusion Matrix: (2x2) Matrix

Predicted Values ($\hat{y}$) \ Actual values (y)	1	0
1	3	2
0	1	1

(Counting 0 1 combination from dataset)

	1	0
1	True Positive (TP)	False positive (FP)
0	False Neg (FN)	True Neg (TN)

	1	0 (Y)
1	TP	FP
0	FN	TN

→ Correct Prediction

$$\text{Model acc} = \frac{TP+TN}{TP+FP+FN+TN}$$

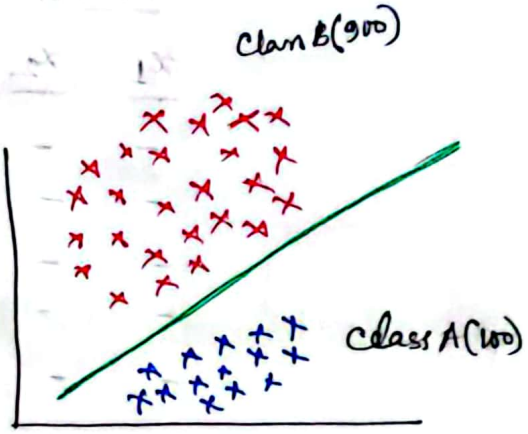
$$= \frac{3+1}{3+2+1+1}$$

$$= 57.1\%$$

But if the dataset is imbalanced ?

Suppose,
1000 datasets

- 900 values → 1
- 100 values → 0



"if dataset is imbalanced", then your model will be a dumb model and suppose it provide → 90% accuracy which is bad

So, in this case, In this scenario, we cannot use the above accuracy calculation Formula.

Out of all the "actual" values, how many are correctly predicted

In this problem scenario, we used "precision Formula".

Precision Formula: $\frac{TP}{TP+FP}$

	1	0 → y
1	TP	FP
0	FN	TN

↓ ŷ

Reducing FP ↓↓

We have to use Precision, when "FP → Important"

Means, when we have to reduce the number of FP, we have to use precision.

TP → True Positive → Actual (y) - (1) and Predicted (ŷ) - (1)

FP → False Positive → Actual (y) - (0) and Predicted (ŷ) - (1)

FN → False Negative → Actual (y) - (1) and Predicted (ŷ) - (0)

TN → True Negative → Actual (y) - (0) and Predicted (ŷ) - (0)

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

	1	0
1	TP	FP
0	FN	TN

We have to use ~~preci~~ recall, when we are reducing FN.

$\frac{TP}{TP+FN}$ } Out of all the "predicted" values, how many are correctly predicted with actual values

Scenario when to use Precision Formula:

Suppose, we are making a spam/not spam email predictore.

(y) Mail → Spam → 1
(ŷ) Model prediction → Spam → 1
} Good scenario

	1	0
1	TP	FP
0	FN	TN

(y) Mail → Not a spam → 0
(ŷ) Model prediction → Spam → 1
} FP important Blunder (Wordt scenario)

(y) Mail → Spam → 1
(ŷ) Model prediction → Not spam → 0
} Bad scenario (Ignorable)

In the above scenario FP is very much important Focus area.
We have to reduce the value of FP as much as possible to get accurate results and to reduce wrong predictions.
In this case, we have to use "Precision Performance Metrics"

Usecase 2: (When to use Recall Formula)

Suppose, we are making a model that predict diabetes/Not Diabetes

	1	0
1	TP	FP
0	FN	TN

(y) Patient → Diabetes → 1
($\hat{y}$) Model → Diabetes → 1
} TP Good Scenario

(y) Patient → Diabetes → 1
($\hat{y}$) Model → No Diabetes → 0
} FN Blunder Most case

(y) Patient → No Diabetes → 0
($\hat{y}$) Model → Diabetes → 1
} FP Bad scenario but not worst

In the above scenario, Recall Formula is more accurate to use Because here we have to reduce the Number of FN, Because FN is considered the most important case here.

That's why we will use Recall Formula in this scenario.

F-Beta Score: (When FP and FN Both are important)

Formula: $(1+\beta^2)\ \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$

If, FP and FN Both are important then, $\beta = 1$

F1 Score ~~Formula~~: $2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$ } called Harmonic Mean

if FP is more important than FN, $\beta = 0.5$ ~~(when FPs)~~

F0.5 Formula = $(1+0.25) * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$

If FN is more important than FP, $\beta = 2$

F2 score Formula = $5 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$