Dt:- 18/4/22
2:30 pm

* Supervised Algorithm :-
  it has Label Data

* Predicting :- Discrete (categories) | class (Yes/No)
  * Binary classification [0, 1]
  * Multi classe | classification [drug x, drug, drug
  Ex :-

#Classification [Algorithms]

→ Logistic Regression
→ KNN [K-Nearest Neighbours]
→ Decision Tree
→ Random Forest
→ Ada Boost
→ Gradient Boost
→ XG Boost
→ SVM [Support Vector Machine]
→ Navies Bayes

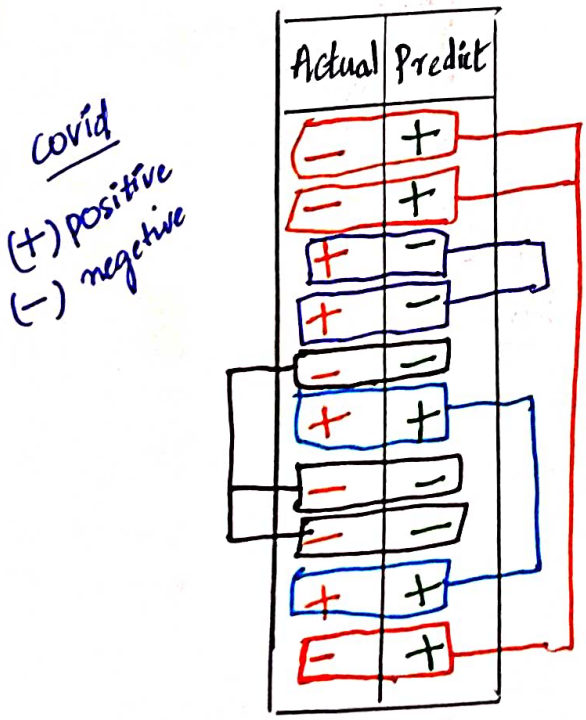
⇒ Classification Use Cases :

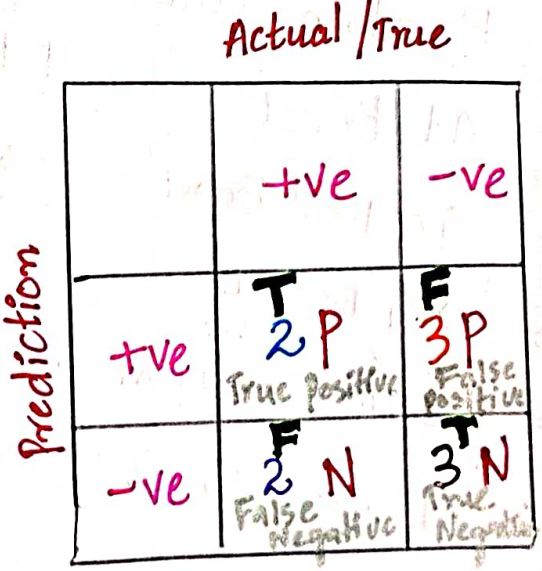
1. Which **Category** a customer belongs to ?
2. Whether Customer **Switches** to another provider / brand ?
3. Whether customer **responds** to a particular advertising Campaign ?

⇒ Evaluation Metrics :

⇒ Confusion Matrix

Ex:- Whether The person is covid: (+) positive
(-) negative.





⇒ T/F - Correct Prediction | incorrect Prediction

⇒ P/N - Predicted Positive | Predicted Negative

* Another Type

Actual \ Predict	+ve	-ve
+ve	TP	FN
-ve	FP	TN

Actual \ Predict	+ve	-ve
+ve	2 [TP]	2 [FN]
-ve	3 [FP]	3 [TN]

* Classification [Accuracy] = $\frac{\text{Correct Predictions}}{\text{total predictions}} = \frac{5}{10}$

* Mis-classification [Error] = $\frac{\text{Incorrect prediction}}{\text{total predictions}} = \frac{5}{10}$

Que :- Which is more important "False positive", "False Negat

Ans :- it is dependent on "Bussiness problem" → FN ↓
FP ↓
Error ↓
Accuray ↑

⇒ Person is Having covid

Actual +ve

Predicted -ve

Always Predicted Comes last (Second Letter)

[FN] More important
Because. if person is Having covid,
You predicted negative, Problem: (i) die
(ii) Spread

F N

⇒ Person is not having covid

Actual -ve

Predicted +ve

[FP]

F P

If person is not having covid,
You predicted covid, (i) quarantine
(ii) Antibotics (medicins)

[FP] More important
But, Person is Havin
cancer, But predict
positive.
(i) Kimotherapy

Que :- In bussiness problem, They said to Focus of "FN" shou

be low?

If client, wants Accuracy and FN

↓ Best

	Actual	
Pred	1 TP	1 FP
	8 FN	TN 90

Accuracy = 91 %.

FN = 8

	Actual	
Pred	4 TP	6 FP
	3 FN	87 TN

Accuracy = 91 %.

False Negative = 3

If client, wants only FN should be low, Not Accuracy

	Actual	
Pred	38 TP	24 FP
	0 FN	38 TN

Accuracy = 76 %

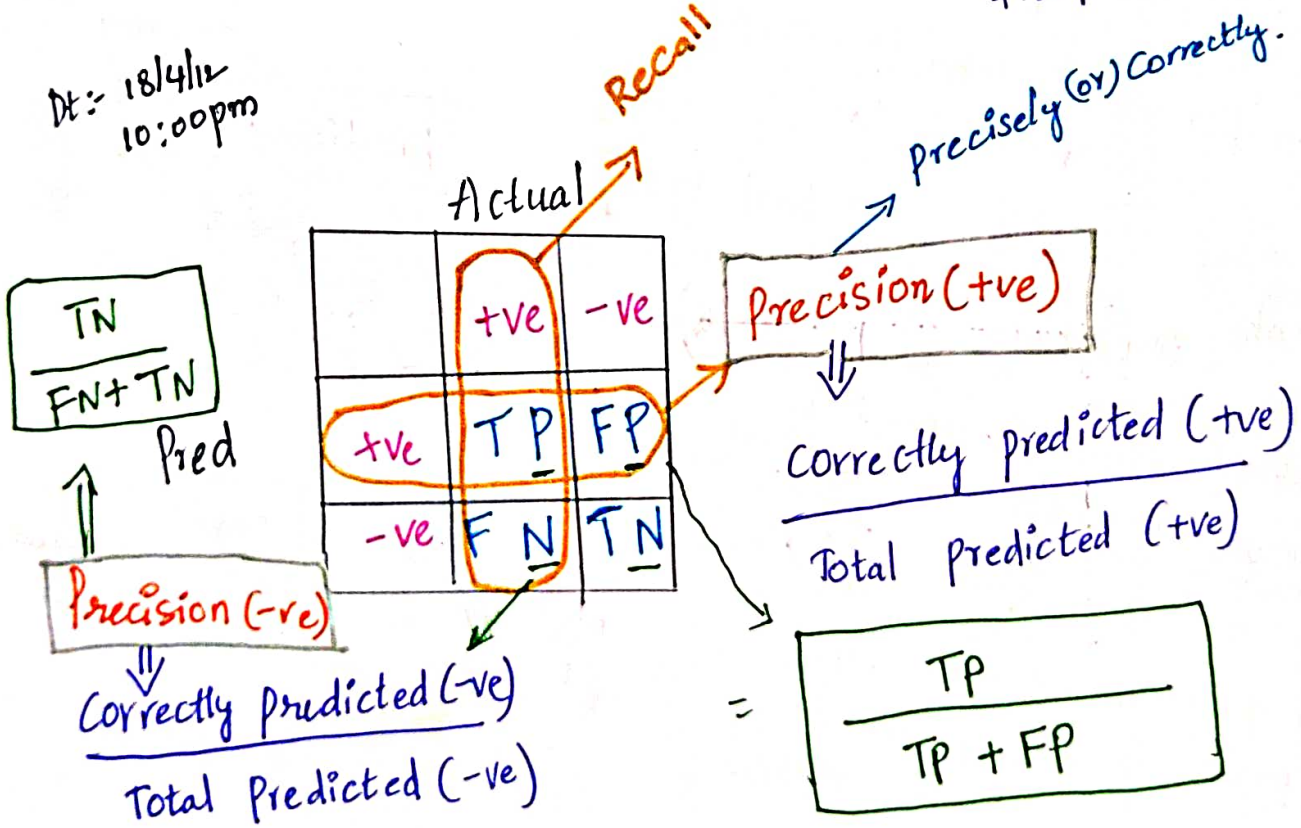
FN = 0

← best

18/4/22

4:30pm.

Dt :- 18/4/22
10:00pm



2) Recall (+ve) = Correctly Predicted (+ve) / total Actual Positive (+ve)

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

* Recall (-ve) = correctly predicted (-ve) / total Actual negative (-ve)

$$\frac{TN}{FP + TN}$$

** Que :- What is difference between Recall and Precision ?

Ans :- Precision :- Comparing with The Predicted Values

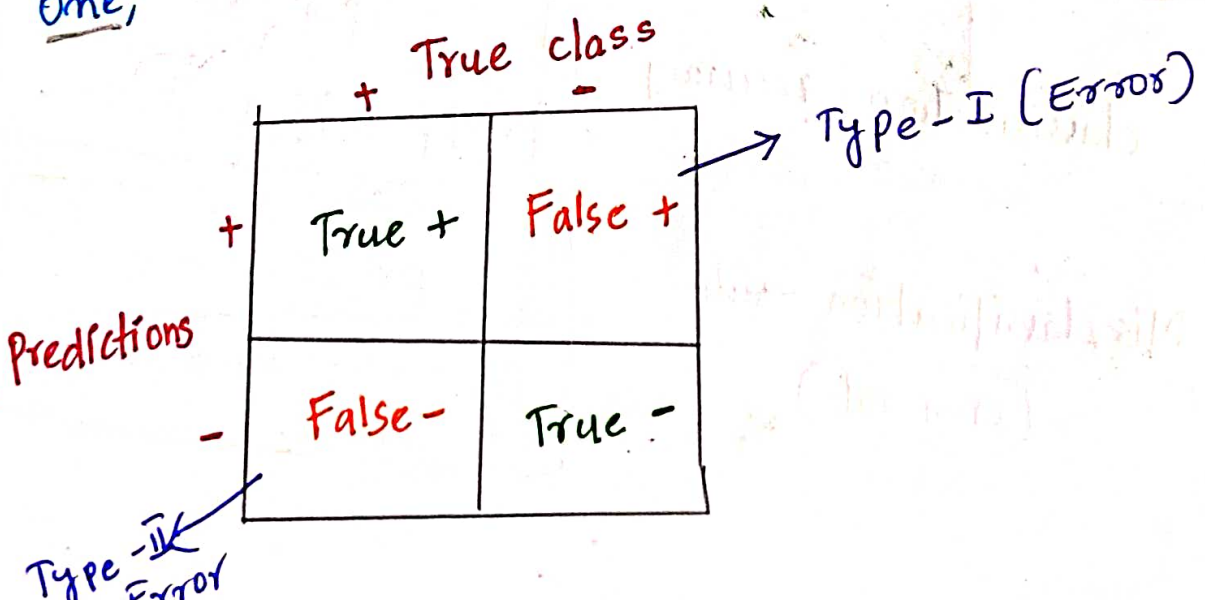
Recall :- Comparing with Actual values.

Preci	Recall
Comparing with Predicted	Actual

* denominator changes, where numerator is same

* denominator instead of prediction, Take actual

One, instead of predicted Positive, Take actual positive



* True positive [TP] :- when classifer predicted True and correct class was True
Ex: They have disease, predicted correctly.

* True Negative [TN] : when model predicted False and correct class was False
Ex:- No disease, But predictive not have disease.

* False positive [FP] [Type-I Error]
classifier predicted True, but correct class was False.
Ex:- Patient did not have disease.

* False Negatives [FN] [Type-II Error]
classifier predicted False, but they actually do have disease.
Ex:- Patient do not have disease.

** Key Performance Indicators [KPI]

classification accuracy = $\frac{TP + TN}{(TP + TN + FP + FN)}$

Misclassification rate (error rate) = $\frac{FP + FN}{(TP + TN + FP + FN)}$

Example

Pred \ Actual	+	−
+	TP = 1	FP = 1
−	FN = 8	TN = 90

classification accuracy = $\frac{TP + TN}{TP + FP + FN + TN}$ = 91 %

precision (+ve) = $\frac{TP}{TP + FP}$ = $\frac{1}{2}$ = 50 %.

Recall (+ve) = $\frac{TP}{TP + FN}$ = $\frac{1}{1+8}$ = $\frac{1}{9}$ = 11 %.

precision (-ve) = $\frac{TN}{FN + TN}$ = $\frac{90}{8+98}$ = $\frac{90}{98}$ = 0.91 %.

Recall (-ve) = $\frac{TN}{FP + TN}$ = $\frac{90}{1+90}$ = $\frac{90}{91}$ = 0.98 %.

* TPR (recall) = True Positive Rate = $\frac{TP}{TP + FN}$ ⇒ Actual (+ve)

* FPR = False Positive Rate = $\frac{FP}{TP + FN}$ ⇒ Actual (+ve)

* TNR = True Negative Rate = $\frac{TN}{TN + FP}$ ⇒ Actual (-ve)

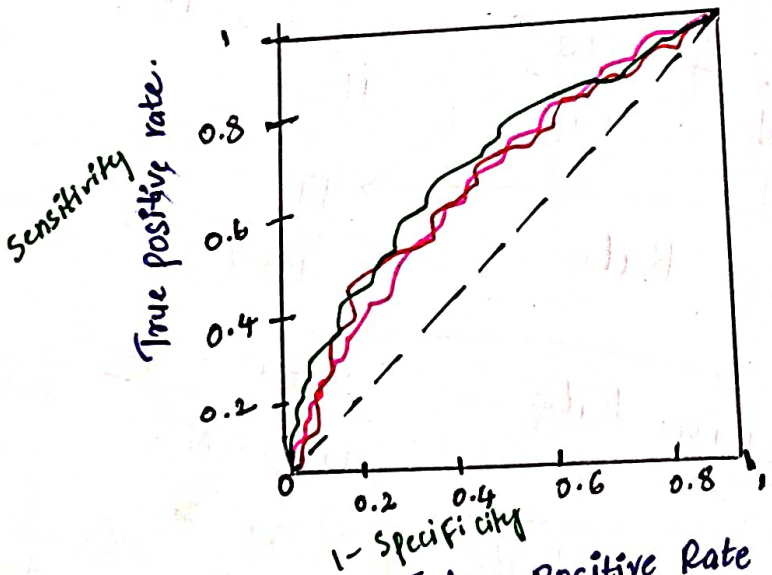
* FNR = False Negative Rate = $\frac{FN}{FP + TN}$ ⇒ Actual (-ve)

* Specificity = TNR ⇒ 1 - FPR

= True Negative Rate = $\frac{\text{True Negatives}}{\text{True negatives + False positives}}$

* Sensitivity ⇒ TPR ⇒ Recall ⇒ $\frac{TP}{TP+FN}$

ROC [Reciver operating characteristic Curve]



False Positive Rate = 1 - specifity = 1 - TNR

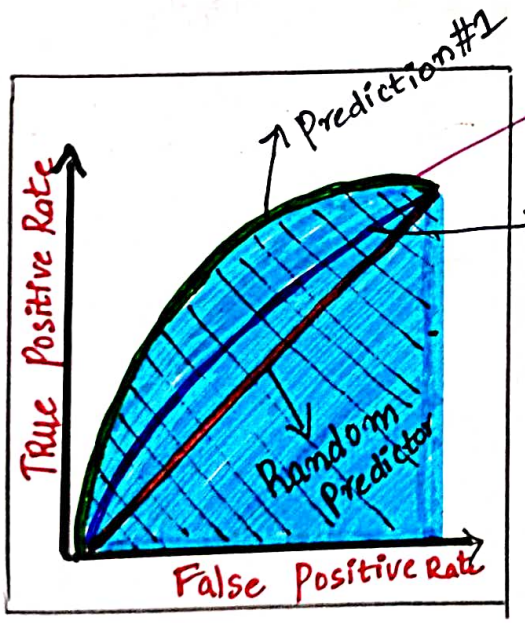
* ROC curve is a metrics that asses model ability to distinguish between binary (0 (or) 1) classes

* Roc curve is created by plotting the true positive rate (TPR) against the False positive Rate [FPR] at various threshold settings

* True positive rate is also known as Sensitivity, recall (or) Probability of detection in Machine Learning.

* False positive rate is also known as the probability of False alarm and can be calculated as [1 - specificity]

AUC [Area Under Curve]



the light Blue area represents area under Curve of receiver operating characteristic [AUROC]

Higher the AUC, the better the model is at predicting

18/04/22
11:30 pm