Performance Metrics (Binary Classification)

Confusion Matrix
- It is a 2 by 2 matrix
- We have actual results from our model and then there is our prediction.
- Based on that we try and find different metrics.

		Actual values	
		1	**0**
Predicted values	**1**	TP	FP ↓
	0	FN ↓	TN

- *Target column needs to be balanced before we check the performance metrics*

Accuracy
- **Formula**= TP+TN/(TP+TN+FP+FN)

Precision
- Out of **all the predicted results** how many are correctly predicted.
- **Formula** = TP/(TP+FP)
- Focus is on **FP to reduce it**
- Ex. mail is not spam but model predicts it to be spam.

Recall or Sensitivity

- Out of **all the actual results** how many are correctly predicted
- **Formula** = TP/(TP+FN)
- Focus is on **FN, to reduce it**
- Ex. mail is spam but model predicts it to be not spam.

F-beta score

- When both precision and recall are important
- **Formula**= $((1+beta^2)*(Pre * Recall))/(beta^2(Pre + Recall))$
- **beta=1** when both FP and FN are important.
- **beta=0.5** when FP is more important than FN (FP>FN)
- **beta=2** in other case.

Specificity

- A test returns a negative result given that the that patient does not have the disease.
- **Formula**=TN/(TN+FP)

ROC and AUC Curve

- ROC is a performance measurement curve for the classification problems.
- To decide what threshold value we need in our model to decide my 1's and 0's by default we take as 0.5 in which anything above 0.5 is 1 and below 0.5 is 0.
- These values should be carefully decided in the model.

y-Actual	ÿ- Model Predicted	ÿ(0)	ÿ(0.2)
1	0.8	1	1
0	0.96	1	1
1	0.4	1	1
1	0.3	1	1