Log Normalisation

→ Handle the Outliers
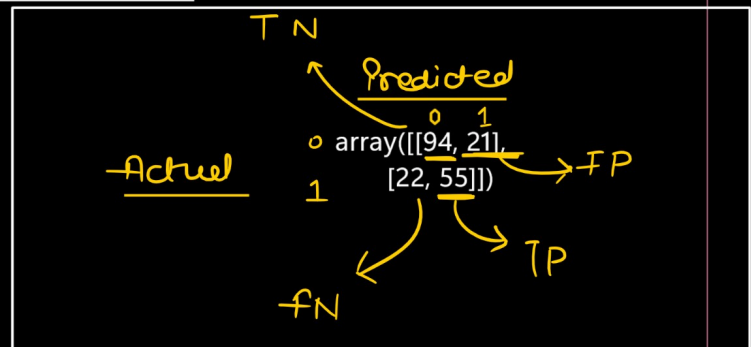
KNN
Algorithm

SMOTE

① Oversampling

Random

② Undersampling

Data

Balanced

$0 \longrightarrow 500$

$1 \longrightarrow 268$ → 500

$0 \longrightarrow 500$ → 268

$1 \longrightarrow 268$

Documentation →

TN

Predicted

     0   1

0 array([[94, 21],

Actual   1   [22, 55]]) → FP

→ TP

fN

$$\text{Recall} \longrightarrow \frac{TP}{TP + FN} = \frac{55}{55 + 22}$$

$$\Rightarrow 0.714$$

ROC-AUC curve

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

Perfect classifier

ROC curve

Better

Worse

Random classifier

True positive rate

False positive rate

Reciever Operating Cherecterstic

↳ across different classification thresholds

—

Task

↳ threshold ≈ 0.2

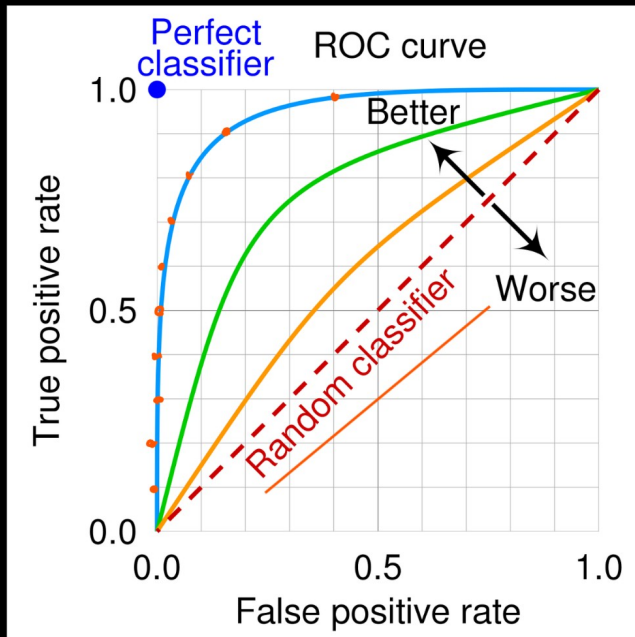↳ What will be the impact & real time When should we go ??

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

↳ threshold ≈ 0.8

↳ What will be the impact & real time When should Actual we go ?? Data

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

Prediction

		0	1
	0	TN	FP
	1	FN	TP

$$Sensitivity = Recall = TPR = \frac{TP}{TP + FN}$$