Data Science Interview Questions

Nemetschek Group SE

A.I. Hub R&D Team

1   Spam Classifier

Based on historical data we observe 20% messages that are "Spam" ("Positives"), i.e. $P(Y = 1)$. We trained a classifier with the following statistical properties:

- True Positive Rate (TPR aka "Recall"): 95%

- False Positive Rate (FPR): 10% $\rightarrow P(\hat{Y} = 1|Y = 0)$

Question: What is the *Precision* of this classifier (in %)?

Consider confusion matrix:

	$\hat{Y} = 1$	$\hat{Y} = 0$
$Y = 1$	TP	FN
$Y = 0$	FP	TN

$Precision = \frac{TP}{TP+FP}$