

Lab 1 Theoretical

	Predicted	
	negative	Positive
actual negative	TN	FP
actual Positive	FN	TP

	Predicted	
	negative	Positive
actual negative	990	10
actual Positive	20	30

$$TN = 990$$

$$FP = 10$$

$$FN = 20$$

$$TP = 30$$

$$\text{Precision} : \frac{TP}{(TP + FP)} = \frac{30}{(30 + 10)} = 0,75$$

$$\text{Recall} : \frac{TP}{(TP + FN)} = \frac{30}{(30 + 20)} = 0,6$$

$$\text{Accuracy} : \frac{(TP + TN)}{(TP + TN + FP + FN)} = \frac{(30 + 990)}{(30 + 990 + 10 + 20)} = 0,9714$$

$$F1 \text{ Score} : 2 \cdot \frac{(\text{Precision} \cdot \text{Recall})}{(\text{Precision} + \text{Recall})} = 0,6667$$

		Predicted	
		no	Yes
actual	no	9000	50
	Yes	100	850

$$TN = 9000$$

$$FP = 50$$

$$FN = 100$$

$$TP = 850$$

$$\text{Precision} : \frac{850}{(850+50)} = 0,944$$

$$\text{Recall} : \frac{850}{(850+100)} = 0,8947...$$

$$\text{Accuracy} : \frac{(850+9000)}{(850+9000+50+100)} = 0,985$$

$$F1 \text{ score} : 2 \cdot \frac{(0,944 \cdot 0,8947)}{(0,944 + 0,8947)} \approx 0,9184$$

Accuracy is not good in cases of imbalanced datasets. High accuracy is easy to get by predicting the dominant class. This ignores the minority class and can be very bad in some cases, such as "heart disease or not".

Precision, Recall and F1 gives more balanced view