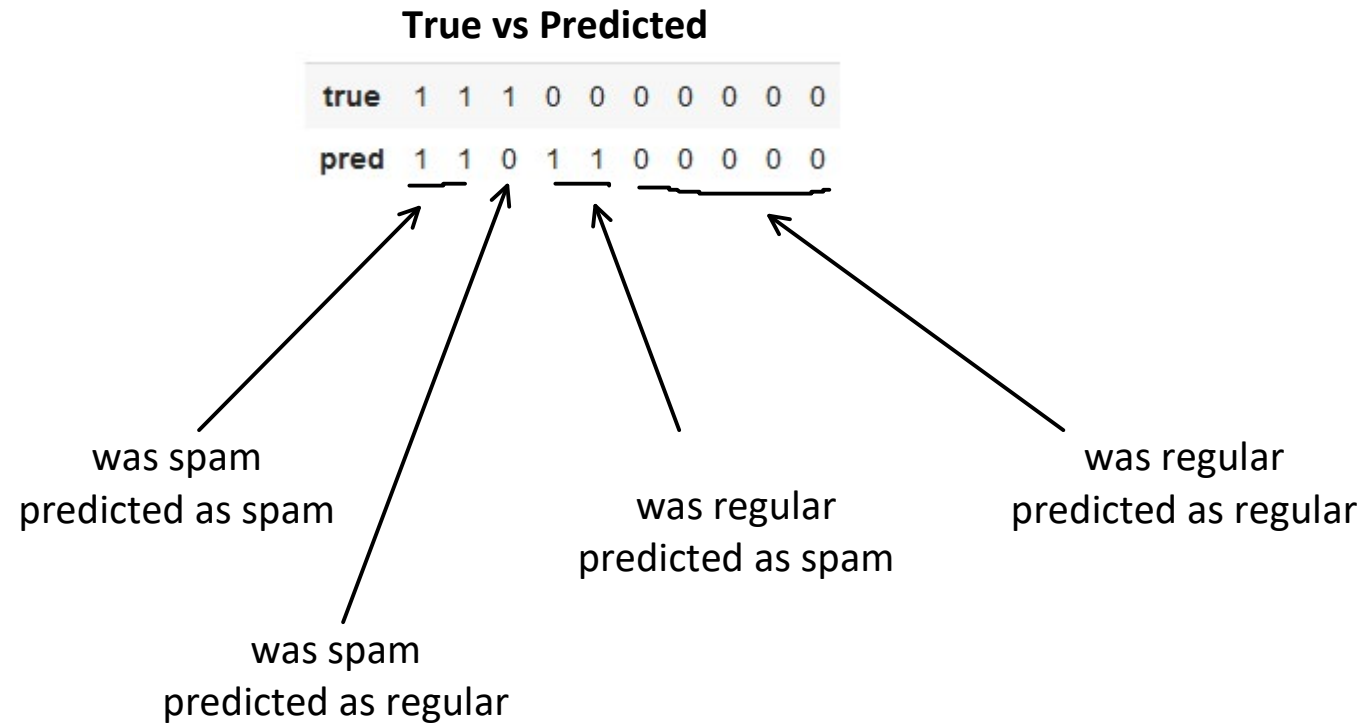


Label	Class
1	Spam Mail (positive)
0	Regular Mail (negative)

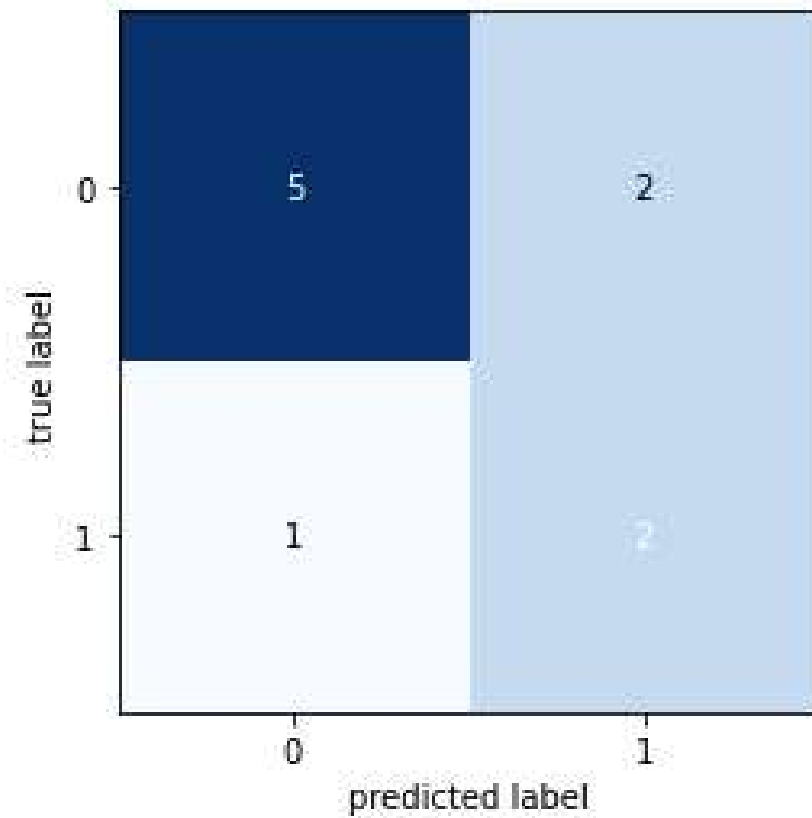
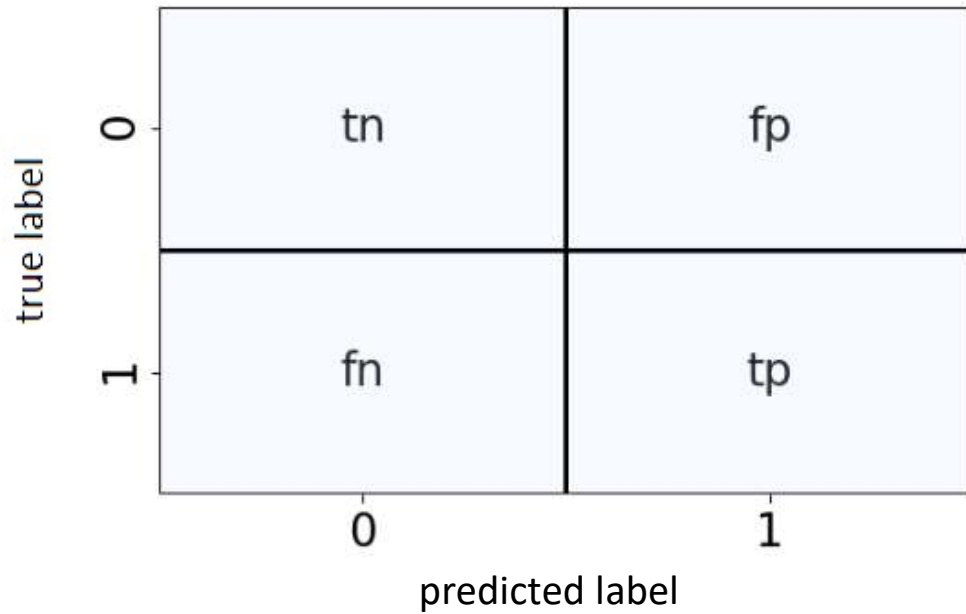


# Confusion Matrix (Two Classes)

Sonntag, 5. Dezember 2021 10:40

true label	predicted label	name	description
positive	positive	<b>tp</b>	true positive predicted
negative	negative	<b>tn</b>	true negative predicted
positive	negative	<b>fn</b>	false negative predicted
negative	positive	<b>fp</b>	false positive predicted

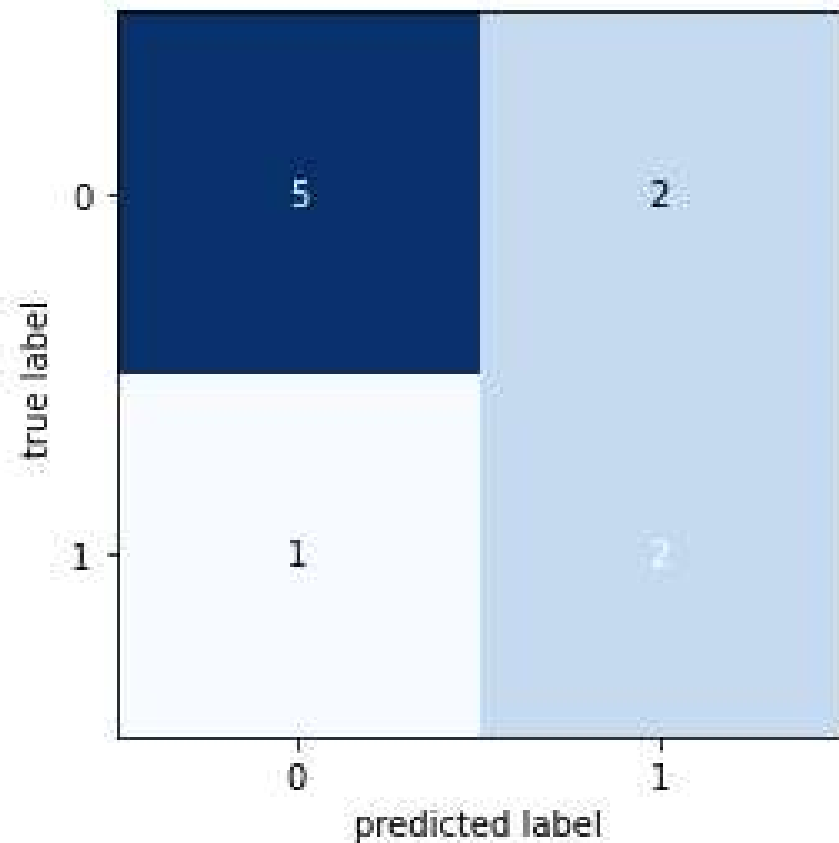
<b>true</b>	1	1	1	0	0	0	0	0	0
<b>pred</b>	1	1	0	1	1	0	0	0	0



Label	Class
1	positive
0	negative

Sonntag, 5. Dezember 2021 11:11

```
true 1 1 1 0 0 0 0 0 0 0
pred 1 1 0 1 1 0 0 0 0 0
```



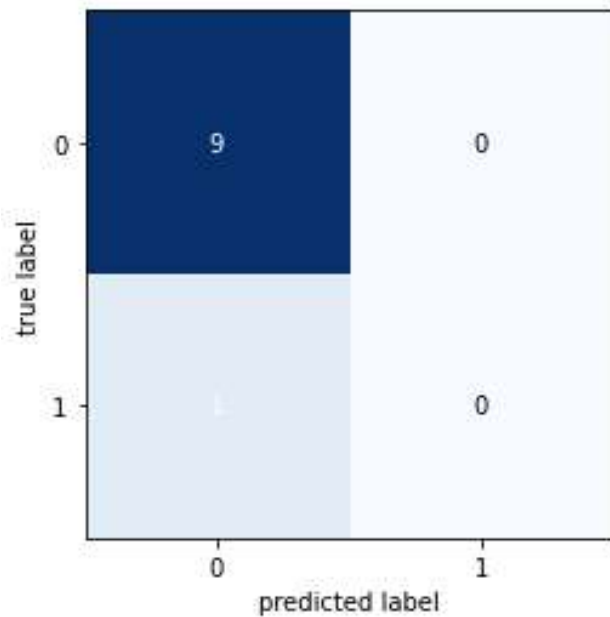
0	tn	fp
1	fn	tp
	0	1

$$\text{accuracy} = \frac{\text{tp} + \text{tn}}{\text{tp} + \text{fp} + \text{tn} + \text{fn}}$$

$$\text{accuracy} = \frac{2 + 5}{2 + 2 + 5 + 1} = 0.7 = 70\%$$

- Assumption: Positive cases are rare
- Dummy classifier: Always predicts *negative*

```
true 1 0 0 0 0 0 0 0 0 0 0  
pred 0 0 0 0 0 0 0 0 0 0 0
```



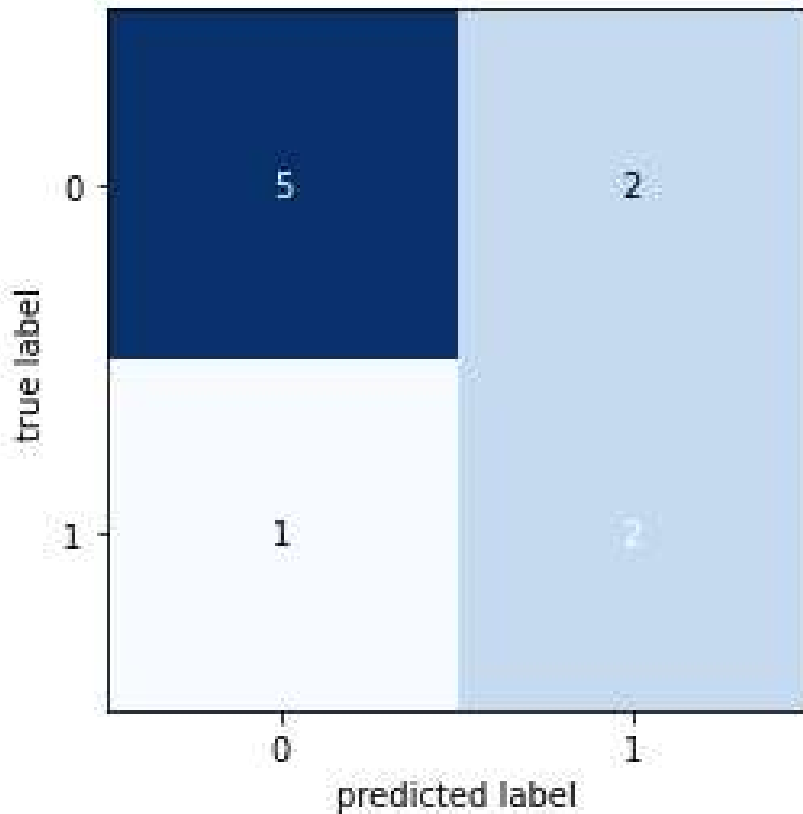
0	tn	fp
1	fn	tp
	0	1

$$\text{accuracy} = \frac{0 + 9}{0 + 0 + 9 + 1} = 0.9 = 90\%$$

but classifier is useless

Fraction of true positive predicted to all positives

```
true 1 1 1 0 0 0 0 0 0 0  
pred 1 1 0 1 1 0 0 0 0 0
```



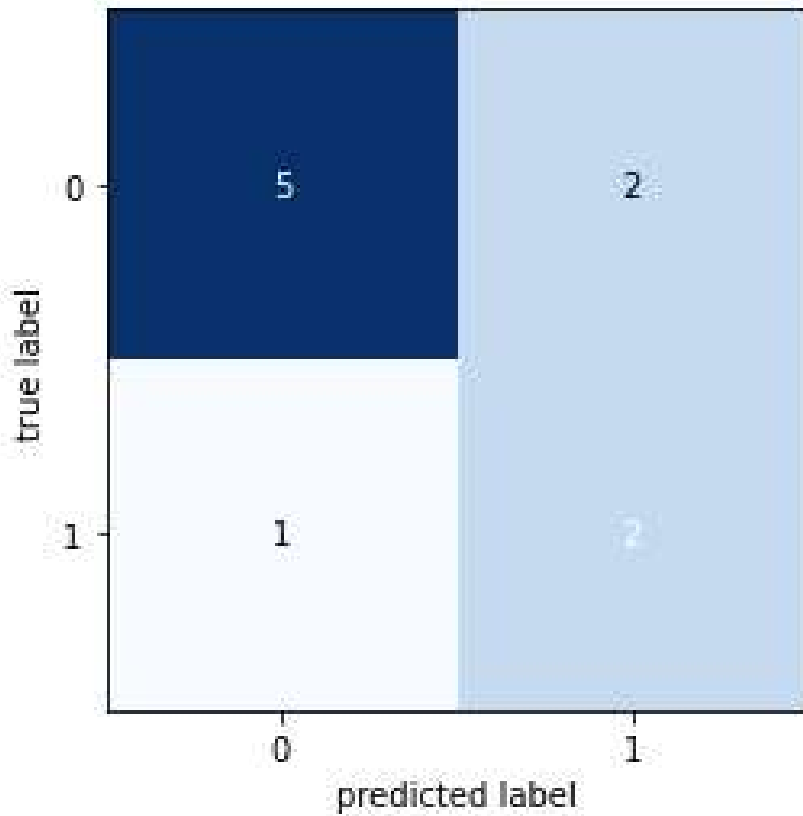
	0	1
0	tn	fp
1	fn	tp

$$\text{recall} = \frac{\text{tp}}{\text{tp} + \text{fn}}$$

$$\text{recall} = \frac{2}{2 + 1} = 0.66 = 66\%$$

Fraction of true positive predicted to all positive predicted

```
true 1 1 1 0 0 0 0 0 0 0  
pred 1 1 0 1 1 0 0 0 0 0
```



0	tn	fp
1	fn	tp
	0	1

$$\text{recall} = \frac{\text{tp}}{\text{tp} + \text{fn}}$$

$$\text{precision} = \frac{2}{2 + 2} = 0.5 = 50\%$$

- Easy to get high recall: just predict positive
  - recall = 100% -> precision bad
- Easy to get high precision: only predict one sample as positive where you are really sure
  - precision = 100% -> recall bad
- Calculate mean between recall and precision
  - Arithmetic mean not a good choice
    - precision = 100%, recall = 1% -> mean(precision, recall) = 50.5%
    - Too good a value for this bad classifier
  - Better: Harmonic mean - F1 value

$$F1 = \frac{2}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}} = \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$