

## Precision and Recall

Let's say the pile has 15 animals:

- 10 Cats
- 5 Dogs

Now, you picked 8 animals from the pile:

- 6 are Cats (Correct selections – "True Positives")
- 2 are Dogs (Mistakes – "False Positives")
- There are still 4 Cats left in the pile (Missed – "False Negatives").

### Precision:

- Out of the 8 animals you picked, 6 are actually cats, so **Precision = 6/8 = 75%**.

### Recall:

- Out of the total 10 cats in the pile, you found 6, so **Recall = 6/10 = 60%**.

### Precision:

- **Definition:** Precision is the proportion of **correctly identified positive results** (True Positives) out of **all the results we labeled as positive**.

Formula:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

### 2. Recall:

- **Definition:** Recall is the proportion of **correctly identified positive results** (True Positives) out of **all actual positives** in the dataset.

- Formula:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

## Example:

Let's say we're trying to classify whether emails are **spam** (positive) or **not spam** (negative). We have a dataset of 20 emails:

- 12 are actually spam (positive cases).
- 8 are not spam (negative cases).

We build a classifier and make the following predictions:

- You predicted that 10 emails are spam.
  - **8 of them are actually spam** (True Positives).
  - **2 of them are not spam** (False Positives).
- We missed **4 spam emails** (False Negatives).

Now, let's calculate **precision** and **recall**:

### Precision:

- Precision focuses on **what you predicted as spam** (positive) and how many of those predictions were actually correct.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} = \frac{8}{8 + 2} = \frac{8}{10} = 0.8 = 80\%$$

- So, your **precision** is **80%**, meaning that **80% of the emails you predicted as spam were actually spam**.

### Recall:

- Recall focuses on **all the actual spam emails** and measures how many of those you correctly identified.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} = \frac{8}{8 + 4} = \frac{8}{12} = 0.6667 = 66.67\%$$

- So, your **recall** is **66.67%**, meaning that **you correctly identified about 67% of all the actual spam emails**.