Lets take an example of email classifying Spam and Not Spam

Precision

- Out of all the emails the model predicted as Spam, how many are correct ?
- Example
  - Out of 10 emails predicted as Spam only 6 are correct and the remaining 4 are important then precision is 6/10
- Precision is important when False positive hurts

Recall

- Out of all the actual Spam mails, how many model is predicted or caught it
- Example
  - There are total 12 spam mails and only 10 model caught or predicted and 2 is missed then recall is 10/12
- Recall is important when false negative hurts

F1-Score

The perfect balance between Precision and Recall.It is harmonic mean calculation

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Macro Average

- Calculates metrics for each class separately
- Then average them equally irrespective of how many examples each class has

Weighted Average

Like macro it calculates metrics pre class
It weighs them by the number of true instances in each class

Accuracy

Overall correctness

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$