Spam Filter Example: TP, FP, FN, TN, Accuracy, Precision, Recall, F1 Score

# Scenario: Testing a Spam Filter

You test 10 emails. Some are actually spam, some are not spam. The AI model tries to predict which ones are spam.

## Email Classification Results

|  |  |  |  |
| --- | --- | --- | --- |
| Email | Is Actually Spam? | Predicted as Spam? | Result Type |
| 1 | Yes | Yes | TP (Correct) |
| 2 | Yes | Yes | TP (Correct) |
| 3 | Yes | No | FN (Missed Spam) |
| 4 | No | Yes | FP (Wrongly flagged) |
| 5 | No | No | TN (Correct) |
| 6 | No | No | TN (Correct) |
| 7 | Yes | No | FN (Missed Spam) |
| 8 | No | Yes | FP (Wrongly flagged) |
| 9 | Yes | Yes | TP (Correct) |
| 10 | No | No | TN (Correct) |

## Count Summary

* TP (True Positive) = 3 → Spam correctly identified
* FP (False Positive) = 2 → Not spam but wrongly flagged
* FN (False Negative) = 2 → Spam missed
* TN (True Negative) = 3 → Not spam and correctly ignored

## Metrics and Definitions

### Accuracy

What It Means: Overall correctness

Formula: (TP + TN) / Total

Value: 60%

### Precision

What It Means: Out of predicted spam, how many were actually spam

Formula: TP / (TP + FP)

Value: 60%

### Recall

What It Means: Out of actual spam, how many were caught

Formula: TP / (TP + FN)

Value: 60%

### F1 Score

What It Means: Balance between precision and recall

Formula: 2 × (P × R) / (P + R)

Value: 60%

## Easy Way to Remember

|  |  |
| --- | --- |
| Term | Think Like... |
| TP | Caught spam correctly ✅ |
| FP | Mistakenly flagged good email ❌ |
| FN | Missed actual spam ❌ |
| TN | Ignored good email correctly ✅ |