Confusion matrix

02 July 2024 12:13

Confusion Matrix

A confusion matrix is a table used to evaluate the performance of a clas
particularly useful in binary and multiclass classification problems. The r
target values with those predicted by the model.

For a binary classification problem, the confusion matrix has four compo

- **True Positives (TP)**: The number of positive instances correctly pr
- **True Negatives (TN)**: The number of negative instances correctly
- **False Positives (FP)**: The number of negative instances incorrectly
  model (Type I error).
- **False Negatives (FN)**: The number of positive instances incorrectly
  model (Type II error).

Here's the layout of a confusion matrix for a binary classification:

	Predicted Positive	**Predicted Negative**
Actual Positive	TP	FN
Actual Negative	FP	TN

Recall Score

Recall (also known as Sensitivity or True Positive Rate) measures the abi
positive instances. It is the ratio of correctly predicted positive observat
Recall is defined as:
Recall=TPTP+FN\text{Recall} = \frac{TP}{TP + FN}Recall=TP+FNTP

Worked Example

Let's assume we have a binary classification problem with the following

- True Positives (TP) = 50
- True Negatives (TN) = 40
- False Positives (FP) = 10

sification algorithm. It is
matrix compares the actual

onents:
edicted by the model.
predicted by the model.
predicted as positive by the

predicted as negative by the

lity of a classifier to find all the
ions to the actual positives.

results from a confusion matrix:

- False Negatives (FN) = 5

The confusion matrix looks like this:

	Predicted Positive	**Predicted Negative**
Actual Positive	50	5
Actual Negative	10	40

Using these values, we can calculate the recall:
Recall=TPTP+FN=5050+5=5055≈0.909\text{Recall} = \frac{TP}{TP + FN} =
\approx 0.909Recall=TP+FNTP=50+550=5550≈0.909

So, the recall score is approximately 0.909, or 90.9%. This means that th
about 90.9% of all actual positive instances.

Importance of Recall

- **High Recall**: Indicates that most of the actual positives are correct
  applications where missing a positive case is costly (e.g., disease d
- **Low Recall**: Suggests that many actual positives are missed by the
  unacceptable in critical applications.

In summary, the confusion matrix provides a detailed breakdown of the
recall score focuses on the model's ability to correctly identify positive i
valuable insights into the performance of a classification algorithm.
Anonymous

= \frac{50}{50 + 5} = \frac{50}{55}

e classifier correctly identifies

ly identified. It is crucial in
letection, fraud detection).
model, which may be

classification results, while the
nstances. Together, they offer