Confusion Matrix

A confusion matrix is a table often used to describe the performance of a classification algorithm, typically a supervised learning one. It allows you to visualize the algorithm's performance and understand where it might be making mistakes. Here are the steps to understand a confusion matrix:

	Predicted Class: Positive	Predicted Class: Negative	
Actual Class: Positive	True Positive (TP)	False Negative (FN) Type II Error	**Sensitivity** $\frac{TP}{(TP+FN)}$
Actual Class: Negative	False Positive (FP) Type I Error	True Negative (TN)	**Specificity** $\frac{TN}{(TN+FP)}$
	Precision $\frac{TP}{(TP+FP)}$	**Negative Predictive Value** $\frac{TN}{(TN+FN)}$	**Accuracy** $\frac{TP+TN}{(TP+TN+FP+FN)}$

1. **Understanding the Basics:**
   - A confusion matrix compares the actual target values with those predicted by the machine learning model.
   - This comparison results in four different outcomes: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).
2. **Components of a Confusion Matrix:**
   - **True Positive (TP):** The cases in which the true class was positive, and the model also predicted positive.
   - **True Negative (TN):** The cases in which the true class was negative, and the model also predicted negative.
   - **False Positive (FP):** The cases in which the true class was negative, but the model predicted positive.
   - **False Negative (FN):** The cases in which the true class was positive, but the model predicted negative.
3. **Calculating Metrics:**
   - **Accuracy:** The ratio of correctly predicted observation to the total observations. Accuracy=$(TP+TN+FP+FN)(TP+TN)$
   - **Precision:** The ratio of correctly predicted positive observations to the total predicted positives. Precision=$(TP+FP)TP$
   - **Recall (Sensitivity):** The ratio of correctly predicted positive observations to all observations in the actual class. Recall=$(TP+FN)TP$
   - **F1 Score:** The weighted average of Precision and Recall. F1 Score=2×(Precision+RecallPrecision×Recall)

https://github.com/BytesOfIntelligences

4. **Interpreting the Confusion Matrix:**
    - High precision and low recall mean that the model is very sure about its positive predictions, but it's missing a lot of actual positives.
    - High recall and low precision means that the model is capturing a lot of actual positives, but it's also making a lot of false positive predictions.
    - A good model should have a balance between precision and recall, and this is where the F1 score comes in handy as it considers both precision and recall.
5. **Use Cases:**
    - Confusion matrices are used in various domains such as medical diagnosis, spam filtering, and more.
    - In medical diagnosis, for example, a false negative (predicting no disease when there is one) can be more serious than a false positive (predicting a disease when there isn't one).

By analyzing the confusion matrix, you can gain insights into where your model might be going wrong and what kind of errors it is making. This can then inform your next steps in terms of model improvement or data collection.