**percentage accuracy:**

The following formula is used to calculate the percentage accuracy of a test.

𝐴 = 100 − [(𝑇𝑣−𝑂𝑣)/𝑇𝑣∗100]

Where A is the percentage accuracy (%)

Tv is the true value or theoretical value

Ov is the observed or measured value

To calculate a percentage accuracy, subtract the observed value from the true value, divide by the true value, multiply by 100, then subtract this result from 100.

Classification accuracy is the total number of correct predictions divided by the total number of predictions made for a dataset.

As a performance measure, accuracy is inappropriate for imbalanced classification problems.

The main reason is that the overwhelming number of examples from the majority class (or classes) will overwhelm the number of examples in the minority class, meaning that even unskillful models can achieve accuracy scores of 90 percent, or 99 percent, depending on how severe the class imbalance happens to be.

**Confusion matrix:**

**| Positive Prediction | Negative Prediction**

**Positive Class** | True Positive (TP) | False Negative (FN)

**Negative Class** | False Positive (FP) | True Negative (TN)

The precision and recall metrics are defined in terms of the cells in the confusion matrix, specifically terms like true positives and false negatives. Precision is a metric that quantifies the number of correct positive predictions made.

**Precision:** (accuracy for the minority class)

Precision quantifies the number of positive class predictions that actually belong to the positive class.It is calculated as the ratio of correctly predicted positive examples divided by the total number of positive examples that were predicted.

Precision = TruePositives / (TruePositives + FalsePositives)

The result is a value between 0.0 for no precision and 1.0 for full or perfect precision.

**Python code:**

# calculate precision python

precision = precision\_score(y\_true, y\_pred, average='binary')

print('Precision: %.3f' % precision)

**Recall:**

Recall quantifies the number of positive class predictions made out of all positive examples in the dataset.

Recall is a metric that quantifies the number of correct positive predictions made out of all positive predictions that could have been made.

Unlike precision that only comments on the correct positive predictions out of all positive predictions, recall provides an indication of missed positive predictions

Recall = TruePositives / (TruePositives + FalseNegatives)

The result is a value between 0.0 for no recall and 1.0 for full or perfect recall.

**Python code:**

# calculate recall in python

recall = recall\_score(y\_true, y\_pred, average='binary')

print('Recall: %.3f' % recall)

**F-Measure:**

F-Measure provides a way to combine both precision and recall into a single measure that captures both properties. Alone, neither precision or recall tells the whole story. We can have excellent precision with terrible recall, or alternately, terrible precision with excellent recall. F-measure provides a way to express both concerns with a single score.

F measure is calculated as follows:

F-Measure = (2 \* Precision \* Recall) / (Precision + Recall)

This is the harmonic mean of the two fractions. This is sometimes called the F-Score or the F1-Score and might be the most common metric used on imbalanced classification problems.

**Python Code:**

# calculate score

score = f1\_score(y\_true, y\_pred, average='binary')

print('F-Measure: %.3f' % score)

F-Measure provides a single score that balances both the concerns of precision and recall in one number. Maximizing precision will minimize the number false positives, whereas maximizing the recall will minimize the number of false negatives

Precision: Appropriate when minimizing false positives is the focus.

Recall: Appropriate when minimizing false negatives is the focus.

Sometimes, we want excellent predictions of the positive class. We want high precision and high recall.

An F-score is the harmonic mean of a system’s precision and recall values. Criticism around the use of F-score values to determine the quality of a predictive system is based on the fact that a moderately high F-score can be the result of an imbalance between precision and recall and, therefore, not tell the whole story. On the other hand, systems at a high level of accuracy struggle to improve precision or recall without negatively impacting the other.

Critical (risk) applications that value information retrieval more than accuracy (i.e., producing a large number of false positives but virtually guaranteeing that all the true positives are found) can adopt a different scoring system called F2 measure, where recall is weighed more heavily. The opposite (precision is weighed more heavily) is achieved by using the F0.5 measure.