Evaluate the classifier

First, we predict the labels for the testing data using the trained classifier:

y\_pred = clf.predict(X\_test)

Next step is to compute accuracy, precision, recall and F1 score:

* First, we need to import from sklearn.metrics:

from sklearn.metrics import accuracy\_score, precision\_score, recall\_score, f1\_score

* Accuracy measures the proportion of correctly classified instances out of the total number of instances. The formula to compute accuracy is as follows:

Accuracy = (Number of correctly classified instances) / (Total number of instances)

Code to compute accuracy:

accuracy = accuracy\_score(y\_test, y\_pred)

print("Accuracy:", accuracy)

* Precision is the ratio of true positive predictions to the total number of positive predictions. It measures the classifier's ability to correctly identify positive instances.

Code to compute accuracy:

precision = precision\_score(y\_test, y\_pred)

print("Precision:", precision)

* Recall (also known as sensitivity or true positive rate) is the ratio of true positive predictions to the total number of actual positive instances. It measures the classifier's ability to correctly capture positive instances.

Code to compute accuracy:

recall = recall\_score(y\_test, y\_pred)

print("Recall:", recall)

* The F1 score is a metric commonly used in binary classification tasks to evaluate the performance of a classifier. It is the harmonic mean of precision and recall. The formula to compute the F1 score is as follows:

F1 score = 2 \* (precision \* recall) / (precision + recall)

Code to compute F1 score:

f1 = f1\_score(y\_test, y\_pred)

print("F1 score:", f1)

The result:

1. Accuracy: The accuracy of 0.8317307692307693 suggests that the classifier correctly predicts the class label for approximately 83.17% of the samples in the testing set.
2. Precision: The precision score of 0.8217821782178217 indicates the proportion of correctly predicted positive instances (audio 2) among all instances predicted as positive. It means that around 82.18% of the predicted audio 2 samples are correct.
3. Recall: The recall score of 0.83 represents the proportion of correctly predicted positive instances (audio 2) among all actual positive instances in the testing set. It means that approximately 83% of the audio 2 samples are correctly identified.
4. F1 score: The F1 score of 0.8258706467661692 is the harmonic mean of precision and recall. It provides a balance between precision and recall and is a useful metric when classes are imbalanced.

These metrics indicate a reasonably good performance of the classifier on the testing data. It appears that the classifier is neither overfitting nor underfitting the data. The performance on the testing set is reasonably close to the performance on the training set. This suggests that the model has learned the patterns in the training data and is able to generalize well to unseen data.