Confusion Matrix for your Multi-Class ML Model

By Sunil Ghimire- “**Be Unique, Be Identifiable, Be You**”

For the sake of simplicity, let’s assume our multi-class classification problem to be a 3-class classification problem. Say we have got three class labels(Target/Outcome) in a dataset, namely 0.1, and 2. A potential uncertainty matrix for these groups is provided below.

		Predicted			
		0	**1**	**2**	Support ↓
Actual	**0**	**34**	13	5	52
	1	0	**52**	0	52
	2	13	0	**33**	46
		47	65	38	Total = 150

Figure 01: Confusion Matrix for Multi-Class Classification

Unlike binary classification, there are no positive or negative classes here. At first, it might be a little difficult to find TP, TN, FP, and FN since there are no positive or negative classes, but it’s actually pretty easy.

A. Accuracy

Accuracy is the most commonly used matrix to evaluate the model which is actually not a clear indicator of the performance.

Accuracy = $$\frac{34+52+33}{150} = 0.793$$

B. Misclassification Rate/ Error

Error tells you what fraction of predictions were incorrect. It is also known as Classification Error.

Misclassification Rate/ Error = $1 - \text{Accuracy} = 1 - 0.793 = 0.207$

C. Precision / Positive Predicted Value

Precision is the percentage of positive instances out of the total predicted positive instances which means precision or positive predicted value means how much model is right when it says it is right.

Precision of Class 0 $= \frac{34}{34 + 0 + 13} = 0.723$

Precision of Class 1 $= \frac{52}{13 + 52 + 0} = 0.8$

Precision of Class 2 $= \frac{33}{5 + 0 + 33} = 0.868$

D. Recall / True Positive Rate / Sensitivity

Recall literally is how many of the true positives were recalled (found), i.e. how many of the correct hits were also found.

Recall of Class 0 $= \frac{34}{34 + 13 + 5} = 0.653$

Recall of Class 1 $= \frac{52}{0 + 52 + 0} = 1$

Recall of Class 2 $= \frac{33}{13 + 0 + 33} = 0.7173$

E. F1-Score

F1- Score is the harmonic mean of the precision and recall which means the higher the value of the f1-score better will be the model. due to the product in the numerator, if one goes low, the final F1 score goes down significantly. So a model does well in F1 score if the positive predicted are actually positives (precision) and doesn't miss out on positives and predicts them negative (recall).

2 * ��$_0$* ��$_0$ 0.653

2 * 0.723 *

F1-Score of Class 0 =-= = 0.6886

��$_0$ + ��$_0$ 2 * 0.6563 2 * 0.8 *

��$_1$* ��$_1$

1

0.7223 +

F1-Score of Class 1 =-= = 0.8888

��$_1$ + ��$_1$ 2 *

��$_2$* ��$_2$ 2 * 0.868 *

0.7170

0.8 + 1

F1-Score of Class 2 =-= = 0.785

0.868 + 0.7170 ��$_2$ + ��$_2$

F. Support

Support is the total number of elements in each predicted class. Here, the support for classes 0, 1, and 2 is 52, 52, and 46.

Hence,

Support for Class 0 = 52

Support for Class 1 = 52

Support for Class 2 = 46

G. Micro F1

This is called a micro-averaged F1-score. It is calculated by considering the total TP, total FP and total FN of the model. It does not consider each class individually, It calculates the metrics globally. So for our example,

Total TP = 34 + 52 + 33 = 119

Total FP = (0 + 13) + (13 + 0) + (5 + 0) = 31

Total FN = (13 + 5) + (0 + 0) + (13+0) = 31

Hence,

$$= \frac{119}{119 + 31} = 0.793$$

$$= \frac{119}{119 + 31} = 0.793$$

Now we can use the regular formula for F1-score and get the Micro F1-score using the above precision and recall.

Micro F1-Score = $\frac{2 * 0.793 * 0.793}{0.793 + 0.793} = 0.785$

As you can see When we are calculating the metrics globally all the measures become equal. Also if you calculate accuracy you will see that,

Precision = Recall = Micro F1 = Accuracy

H. Macro F1

This is macro-averaged F1-score. It calculates metrics for each class individually and then takes unweighted mean of the measures. As we have seen from figure "Precision, Recall and F1-score for Each Class",

Precision	Recall F1-Score
Class 0 Precision = 0.723	Class 0 Recall= 0.653 Class 0 F1-Score = 0.686
Class 1 Precision = 0.8	Class 1 Recall=1 Class 1 F1-Score = 0.8888
Class 2 Precision = 0.868	Class 2 Recall= 0.7173 Class 2 F1-Score = 0.785

Hence,

Macro Average for precision = $\frac{0.723 + 0.8 + 0.868}{3} = 0.797$

Macro Average for Recall = $\frac{0.653 + 1 + 0.7173}{3} = 0.7901$

$$\frac{0.686 + 0.8888 + 0.785}{}$$

Macro Average for F1-Score =-$= 0.7866$ 3

I. Weighted Average

Weighted Average is the method of calculating a kind of arithmetic mean of a set of numbers in which some elements of the set have greater (weight) value than others. Unlike Macro F1, it takes a weighted mean of the measures. The weights for each class are the total number of samples of that class.

$$0.723 * 47 + 0.8 * 65 + 0.868 * 38$$

Weighted Average for precision =-$= 0.7931$ 150

$$0.653 * 52 + 1 * 52 + 0.7173 * 46$$

Weighted Average for Recall =-$= 0.79301$ 150

2 * ���������h������������������

*���������h����������������������

Weighted Average for F1-Score =-$= 0.79305$ 150

Finally, let's look generated Confusion matrix using Python's Scikit-Learn

In [28]:

```
from sklearn.metrics import classification_report
print(classification_report(y_right, y_predicted))
```

Out [28]:

```
              precision    recall  f1-score   support

           0       0.72      0.65      0.69        52
           1       0.80      1.00      0.89        52
           2       0.87      0.72      0.79        46

    accuracy                           0.79       150
   macro avg       0.80      0.79      0.79       150
weighted avg       0.79      0.79      0.79       150
```

Figure 02: Final Classification Report

50 Questions on Statistics & Machine Learning – Can you answer?

I have tried to summarize detailed [50 tricky and insightful questions on statistics and machine learning] discussion that should benefit you in two ways:

a. Evaluating what you do not know and what you know.

b. Knowing and addressing these similar questions in job interviews help you to generalize your skills.

Figure 03: 50 tricky and insightful questions on statistics and machine learning

The purpose of this article is not to clarify machine learning but to clarify the essential principles that recruiters frequently ask questions in job interviews. Yes, those mentioned questions come to the rescue if you are planning to make your career in Machine Learning and Statistics.

☺ Thanks for your time ☺

What do you think of this "**Confusion Matrix for your Multi-Class ML Model**"? (Appreciation, Suggestions, and Questions are highly appreciated).